

Syddansk Universitet

Measuring regional movement in the lumbar spine

Mieritz, Rune Mygind

Publication date:
2013

Document version
Submitted manuscript

Citation for pulished version (APA):

Mieritz, R. M. (2013). Measuring regional movement in the lumbar spine: Reliability and change in chronic low back pain patients. Syddansk Universitet. Det Sundhedsvidenskabelige Fakultet.

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Measuring regional movement in the lumbar spine

Reliability and change in chronic low back pain patients



Ph.D. Thesis submitted to:

The Institute of Sports Science and Clinical Biomechanics,

Faculty of Health Sciences,

University of Southern Denmark.

February 2013

Rune Mygind Mieritz

Table of Contents

Preface	4
Supervisors	4
List of Manuscripts.....	5
Acknowledgement.....	6
Abbreviations	8
List of Tables and Figures	9
Thesis at a glance.....	10
Introduction and background.....	11
General introduction.....	11
Low back pain definition.....	12
The biopsychosocial model	13
Prevalence, course, impact and prognostic factors of low back pain	14
Management of low back pain.....	16
Manual therapy	17
Self-care interventions and education.....	19
Exercise therapy	20
Physiology of pain.....	21
Mechanisms linking movements and pain.....	22
Biomechanical measurements of the lumbar region.....	23
Lumbar kinematics and patient-rated outcomes	25
Reliability and validity	27
Aims and hypothesis of the thesis	30
Specific objectives and hypothesis:	30

Method.....	31
Summary of methods and materials	31
The systematic review	31
Inclusion and exclusion criteria.....	31
Data collection and management.....	31
Analysis and synthesis.....	32
The cohort studies.....	33
Experimental protocol and examiners	33
Study sample	33
Inclusion and exclusion criteria.....	34
Randomization and blinding.....	34
Interventions	35
Data collection and variables of interest.....	36
Measurement protocol	37
Management of kinematic data.....	41
Statistical analysis	44
Shared statistical analysis for Manuscripts II, III and IV	44
Specific statistical analysis for the reliability study on chronic LBP patients (Manuscript II)	45
Specific statistics for analysis of lumbar motion changes (Manuscript III)	46
Specific statistics for analysis of the association between regional lumbar motion and patient-rated outcomes (Manuscript IV).....	46
Summary of results.....	47
Results of the literature review	47
Included articles.....	47
Descriptive items	48
Quality assessment	49
Results of the cohort studies.....	51
Study sample	51
Reliability and measurement error	53
Regional lumbar motion changes	56
Association between change scores in regional lumbar motion and patient-rated outcomes.....	58

Discussion	61
Discussion of main results	61
The quality of reporting in studies on reliability of 3D regional lumbar motion measurement systems	61
The reliability and measurement errors of regional lumbar motion in chronic LBP patients	62
Regional lumbar motion change following treatment interventions	64
Relationships between changes in regional lumbar motion and changes in patient-rated outcomes	65
Methodological considerations	69
Systematic literature review	69
Cohort studies	69
Lumbar motion measurement test procedures	71
Conclusions	77
Perspectives	78
Clinical implications	78
Research implications	79
Summary in English	80
Summary in Danish	82
References	84

Preface

Supervisors

Professor Jan Hartvigsen, DC, PhD (main supervisor)
Institute of Sport Science and Clinical Biomechanics
Part of the Clinical Locomotion Network
University of Southern Denmark, Odense, Denmark
And
Nordic Institute of Chiropractic and Clinical Biomechanics, Odense, Denmark
Part of the Clinical Locomotion Network

Professor Gert Bronfort DC, PhD (project supervisor)
Northwestern Health Sciences University, Minneapolis, MN, USA
And
Nordic Institute of Chiropractic and Clinical Biomechanics, Odense, Denmark
Part of the Clinical Locomotion Network

Professor Per Aagaard, PhD
Institute of Sport Science and Clinical Biomechanics
Faculty of Health Sciences
University of Southern Denmark, Odense, Denmark

Professor Alan Breen, DC, PhD
Institute for Musculoskeletal Research & Clinical Implementation,
Anglo-European College of Chiropractic, Bournemouth, England.

Associate Professor Greg Kawchuk, DC, PhD
Canada Research Chair in Spinal Function,
University of Alberta, Edmonton, Canada.

List of Manuscripts

The thesis is based on the following manuscripts:

Manuscript I

Mieritz R.M., Bronfort G., Kawchuk G., Breen A., Hartvigsen J. Reliability and measurement error of 3-dimensional regional lumbar motion measures: a systematic review. *J Manipulative Physiol Ther.* 2012;35:645-56.

Manuscript II

Mieritz R.M., Bronfort G., Jakobsen M.D., Aagaard P., Hartvigsen J. Reliability and measurement error of regional lumbar motion parameters in 220 chronic low back pain patients. *Submitted.*

Manuscript III

Mieritz R.M., Hartvigsen J., Boyle E. Jakobsen M.D., Aagaard P., Bronfort G. Regional lumbar motion changes in chronic low back pain patients. *Submitted.*

Manuscript IV

Mieritz R.M. , Bronfort G., Hartvigsen J. Regional lumbar motion and patient-rated outcomes. *Manuscript.*

Acknowledgement

This PhD thesis was accomplished as part of a collaborative study between *The Institute of Sport Science and Clinical Biomechanics at The University of Southern Denmark, Odense, Denmark* and *The Northwestern Health Sciences University, Minneapolis, MN, USA*. The study was financially supported by grants from *The Foundation for Chiropractic Research and Postgraduate Education; The Danish Agency for Science, Technology and Innovation; and The University of Southern Denmark*.

First and foremost, I sincerely wish to thank my supervisor *Jan Hartvigsen*, for your high level guidance, positive attitude and endless support during the entire project. I am truly grateful that you had confidence in me from day one.

Secondly, I would like to thank my project-supervisor *Gert Bronfort* for your data collaboration, hospitality and support during the process. Thirdly, I would like to thank my co-supervisors: *Per Aagaard*, for excellent guidance and manuscript assistance, and *Alan Breen* and *Greg Kawchuk* for guidance and helpful criticism and feedback. Furthermore, I would like to extend special thanks to *Markus Due Jakobsen* for excellent programming assistance and support and to *John Rasmussen* for useful discussions and guidance in the development of analytical models for analyzing the three-dimensional data.

I also gratefully acknowledge *Roni Evans* for your hospitality and assistance with manuscript preparation and *Johan Wallin* for expert research librarian assistance in relation to the literature review. I would like to thank *Erik Poulsen* for practical help with the thesis and many good conversations, *Per Kjær* for interesting discussions over the coffee machine, *Henrik Lauridsen* for expert assistance in clinimetric issues, *Eleanor Boyle* for statistical assistance and enlightening discussions, *Ed Owens* for helping me out with the data extraction, *Suzanne Capell* and *Peter Kent* for language editing and educative discussion, and *Raymond Ostelo* and *Werner Vach* for high level clinimetric and statistical guidance.

I would also like to thank my colleagues at The Institute of Sport Science and Clinical Biomechanics at The University of Southern Denmark for their moral support as well as all the participating patients who made this project possible. Thanks also to the staff members who have been a part of this project at Northwestern Health Sciences University. I am grateful to the many people who have put a lot of work into this project.

I would like to thank my father *Erik Mygind* for constructive discussions about the thesis and to the rest of *my family* for your moral support.

Finally, and most importantly, I would like to express my gratitude to my wife *Hanne* for your endless support, tolerance and understanding during the project.

Rune Mygind Mieritz, Odense, 2013

Abbreviations

BMI	Body mass index
CI	Confidence interval
EIH	Exercise-induced hypoalgesia
HEA	Home exercise and advice
ICC	Intraclass correlation coefficient
LBP	Low back pain
LOA	Limits of agreement
RMDQ	Roland Morris Disability Questionnaire
ROM	Range of motion
SD	Standard deviation
SEM	Standard error of measurement
SET	Supervised exercise therapy
SMT	Spinal manipulative therapy
CAM	Complementary and alternative medicine
2D	Two-dimensional
3D	Three-dimensional

List of Tables and Figures

List of Tables

1. Thesis at a glance
2. Overview of all included studies and its individual quality ratings
3. Patient characteristics
4. Reliability and measurement error of lumbar spinal motion parameters
5. Spinal motion reliability and measurement error for LBP patients divided into subgroups
6. Lumbar motion characteristics at baseline and 12-week follow up by treatment group
7. Regional lumbar motion changes vs. clinically relevant changes in all included patients

List of Figures

1. Flow diagram illustrating the study design
2. Flow diagram illustrating of participants available for test-retest analysis and intervention studies
3. Illustrating the CA6000 Motion Analyzer attached on a person in neutral position and the computer used
4. The CA6000 Motion Analyzer with a person in neutral, extension and flexion positions
5. Illustrating a sequence of pictures of the CA6000 Motion Analyzer attached on a person performing circumduction motion
6. Regional lumbar extension and flexion motion displayed in degrees and velocity at a single representative session
7. Regional lumbar extension and flexion motion obtained in a single representative subject, displayed as a position (degrees) - speed (velocity) phase-plot
8. Regional lumbar circumduction motion in a typical patient before and after treatment
9. Illustration of a Bland-Altman plot
10. Flow diagram illustrating study selection
11. Overview of results of the quality assessment in each domain
12. Lumbar motion percentage changes between baseline and 12-week follow up by motion parameters and treatment groups
13. Graphs illustrating examples of extension/flexion motion trials
14. Graphs illustrating sagittal plane movement data from three trials of a circumduction test procedure

Table 1

Thesis at a glance

Paper	Design	Research question	Materials/ participants	Methods	Conclusions
I	Systematic literature review	What is the quality of reporting in studies dealing with reliability of 3D regional lumbar motion measurement systems and what are the reliability scores obtained by these instruments?	15 articles were included	Electronic searches in Pubmed, CINAHL, Embase and Mantis databases. A detailed checklist was developed consisting of descriptive items divided into four domains: study population, testing circumstances, equipment and data analysis/presentation. The descriptive items were used as a foundation for the quality assessment reflecting the reporting level of the included papers.	Overall the level of reporting was incomplete in several domains downgrading the quality of reporting in general. Acceptable reliability coefficients were generally reported indicating that these may be used for research purposes at the group level, but it is uncertain if any can be used at the individual patient level.
II	Test-retest study	What is the reliability and measurement error of regional lumbar motion in the sagittal plane and do they differ between subgroups of patients?	220 chronic LBP patients	Regional lumbar motion data sampled using a six-degrees-of-freedom instrumented spatial linkage system. ICC _(1,1) and limits of agreement were used for statistical analysis.	Sagittal plane regional lumbar motion data from chronic LBP patients may be sufficiently reliable in measurements on groups of patients, but appear unusable at the individual patient level because of the size of the measurement error. Reliability and measurement error vary substantially between subgroups of patients.
III	Secondary analysis of data from an RCT.	Does regional lumbar motion change following exercise therapy or SMT over a 12-week period?	199 chronic LBP patients	Regional lumbar motion data sampled using a six-degrees-of-freedom instrumented spatial linkage system. Non-parametric statistics used.	Regional lumbar motion changes can occur in chronic LBP patients over a 12-week period and these changes may be associated with type of treatment received.
IV	Secondary analysis of data from an RCT.	What are the relationships between changes in regional lumbar motion and changes in the patient-rated outcomes of back pain and RMDQ?	199 chronic LBP patients	Regional lumbar motion data sampled using a six-degrees-of-freedom instrumented spatial linkage system. Correlations and comparison between patients who improved by a clinically relevant amount versus no change.	Overall changes in regional lumbar motion were poorly correlated with patient-rated outcomes in the form of RMDQ and back pain. Regional lumbar motion changes in the subgroups with back pain only or receiving SMT had the best association with patient-rated outcomes of pain and RMDQ.
3D= Three Dimensional, LBP = Low back pain, ICC _(1,1) = Intraclass correlation coefficient, RCT = Randomized clinical trial, SMT = Spinal manipulative therapy, RMDQ= Roland Morris Disability Questionnaire					

Introduction and background

General introduction

This thesis is based on four scientific manuscripts with a strong focus on non-invasive functional measurable elements of regional lumbar motion. The reader will find a brief overview in Table 1 (Thesis at a glance). Manuscripts II, III and IV are based on a collaboration between *The Institute of Sport Science and Clinical Biomechanics at The University of Southern Denmark, Odense, Denmark* and *The Northwestern Health Sciences University, Minneapolis, MN, USA*. The design and principal data collection used for these three studies was performed between 2001 and 2004 by the Minneapolis group under the direction of Professor Gert Bronfort. The study was designed as a mixed-methods approach based on a prospective, observer-blinded, parallel-group, randomized clinical trial. The trial was conducted at the Wolfe-Harris Center for Clinical Studies at the Northwestern Health Sciences University in Bloomington, MN, USA (Figure 1).

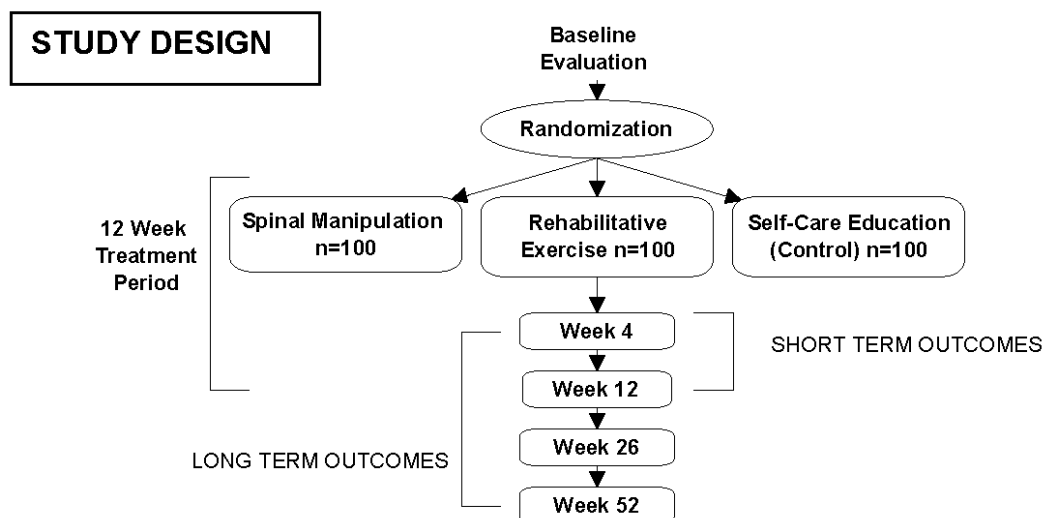


Figure 1: Flow diagram illustrating the study design of the RTC on which the current thesis (Manuscripts II, III and IV) are based.

The primary aim of the study was to examine the relative efficacy of the three interventions (spinal manipulation therapy (SMT), supervised exercise therapy (SET) and home exercise and advice (HEA)) in terms of patient-rated outcomes in the short term (after 12 weeks) and long term (after 52 weeks) for non-acute low back pain (LBP). Secondary aims were to: 1) examine the short- and long-term relative cost-effectiveness and cost utility of the three treatments; 2) assess if there were clinically important treatment group differences between pre-specified subgroups of LBP patients, the subgroups being based on the duration of their current episode of pain (6 to 12 weeks versus more than 12 weeks) and radiating leg pain (present or absent); 3) evaluate if there were

treatment group differences in objective lumbar spine function (range of motion (ROM), strength and endurance) after 12 weeks of treatment and if changes in lumbar function were associated with changes in patient-rated short- and long-term outcomes; 4) identify if baseline demographic or clinical variables could predict short- or long-term outcome; and 5) describe patients' interpretations and perceptions of outcome measures used in clinical trials. This thesis addresses two of the secondary aims (numbers 2 and 3). A paper reporting the results relative to the primary aim has been published recently [1].

In order to have the best possible foundation to deal with the complex regional lumbar (from S1 to T7 spinous process) motion data, experts (national and international) known by the main supervisor were contacted. Based on the guidance of this skilled research group, this project was established.

In the following section, LBP is defined and the biopsychosocial model presented, prior to providing a more detailed background description of the prevalence, course and impact of LBP, its associated prognostic factors and management, physiology of pain and mechanisms linking movements and pain, as well as measurements of the lumbar region and finally, the concepts of reliability and validity.

Low back pain definition

Conceptually, LBP is a symptom, not a disease. The lower back is commonly defined as the area between the bottom of the rib cage and the buttock creases [2] and typically LBP is defined as pain localized within this area. Non-specific LBP is defined as soreness, tension, and/or stiffness in the lower back region for which it is not possible to identify a specific cause [2] and has been estimated to represent approximately 85-90% of the patients suffering from LBP [3;4]. Although non-specific LBP may be considered as one condition, it probably covers several subgroups [5-7].

Pain radiating into the leg(s) can accompany LBP either because it is referred by structures in the spine or because of irritation or compression of nerve roots. LBP may be further classified relative to the duration of the pain (acute, sub-acute and chronic) [8]. The exact timeframe for when the condition becomes chronic is not consistently defined but it usually refers to pain that has lasted longer than 6 to 12 weeks [9-11]. In general, the longer the LBP lasts, the more difficult it is to treat [12].

Pain and symptoms from the low back can arise from spinal structures such as the intervertebral joints and discs, muscles and ligaments although the specific painful structure in the individual patient may be difficult to identify [13]. Many theories have evolved concerning the etiology of LBP but in spite of considerable scientific effort, the mechanisms remain largely unknown [14;15]. Patients with radiculopathy and more serious causes of back pain (e.g. cancer) account for approximately 10% of those with LBP, with the remaining patients' conditions labeled as 'non-specific LBP' or some equivalent term [4]. Therefore, for the majority of LBP patients, a specific

cause of the pain cannot be found and as a consequence, we cannot apply a treatment specifically targeted at the pain source or causal mechanism [12].

“There has always been back pain.” [16]

In the early 1900s, the sacroiliac joint was the main structure thought to induce back pain and manipulative and orthopedic attempts to treat backache by targeting this joint have been described [16]. Later in 1933, the ‘facet syndrome’ was described by Ghormley [17] and the following year Mixter and Bar [18] described the herniated nucleus material as the cause of sciatic pain by compression of the nerve roots. Since then, the primary focus when searching for the etiology of LBP has been on the vertebral discs and surrounding structures. The search for the etiology of LBP still goes on; however there seems to be a growing understanding that the problem is multifactorial and involves more than the disc and the other spinal structures [12].

The biopsychosocial model

The lumbar spine is a complex structure with multiple degrees of freedom, capable of complex movements in all three dimensions. From a biological perspective, the complexity is evident, however interactions between the individual and his/her social and physical environments bring even more complexity to the condition. The biopsychosocial model [19] has helped to illustrate how evolving societal norms influence the individual’s response to LBP. In addition, this model has made it clear that, from a clinical perspective, we must consider the whole patient including his/her environments and not just his/her spine especially where chronicity has developed. Therefore, for the past 3 decades there has been a shift from biologically-oriented to more psychosocially-oriented research and clinical practice in LBP.

The psychosocial factors have been shown to be important factors although the effects of the interventions directed towards them have been shown to be minor [20].

“Despite the large number of studies in the last 10 years, knowledge does not seem to have progressed and moreover, it has not been translated into improved prognosis.” [20]

The enhanced research attention to the psychosocial domain has possibly been at the expense of the biological aspects and, as a consequence, the knowledge of the underlying pathology has advanced little over the last 20 years.

In order to develop a better rationale for choosing the most effective treatment for individual patients, it has been argued that we must also enhance our understanding of the biological component of the biopsychosocial model [4].

For these reasons, some researchers have expressed concern at the lack of research in the biological aspects and requested a more equal focus on all parts of the biopsychosocial model [4]. Therefore, although acknowledging the multidimensional complexity of LBP, this thesis focuses on the biological component of the model; more specifically, the functional measurable elements of spinal motion in the lumbar region.

Prevalence, course, impact and prognostic factors of low back pain

“There is no evidence that the prevalence of back pain has increased over the last 50 years; what has changed is the way individuals, the medical community, and society have responded to back pain.” [21]

LBP is a common and costly condition. Back pain starts early in life [22] and continues to be a common and bothersome complaint into old age [23]. The prevalence of spinal pain has been reported to be fairly evenly spread over all ages, with no obvious increase in prevalence in the elderly although their pain has been reported to be of longer duration [24].

It has been estimated that approximately 60 to 85% of the population will suffer from LBP at some point in their lives [25]. A Danish survey undertaken in 2007 showed that approximately 30% of the population had pain or discomfort in the back or lower back in the previous 14-day period and 15% reported having a back disorder at that time [26]. In a study of Danish twins, 57% reported to have had LBP at some time in their life and 43% had experienced LBP during the previous year [24]. The prevalence of LBP has been reported to be more common in women in all regions of the spine [24;26], with pain reportedly being more frequent and associated with longer sick-leave [27]. LBP compared to neck pain and mid back pain has a greater impact on daily living and LBP more often results in some kind of consequence, e.g. sick-leave, than complaints from other spinal regions [27]. The incidence of LBP appears to be socially skewed where people with more education or higher socio-economic status have lower occurrence [26].

In the acute stage, the prognosis is good i.e. 80% to 90% of people will improve considerably within 6 to 8 weeks [19;28;29]. Approximately 3% to 10% of people with acute back pain have been estimated to develop persistent LBP [9;30;31]. The socio-economic burden of this relatively small group with chronic LBP significantly exceeds that of acute LBP; less than 5% of people who sustain a LBP episode each year account for

75% of the total costs [9;31;32]. Therefore, much effort has been devoted to identifying risk factors for developing persistent or chronic LBP at an early stage in order to reduce costs and increase clinical effectiveness [30;33]. Although there is a wide variation in the cost estimates among different studies, there is no doubt that it is a heavy economic burden for western societies. A recent Danish cost analysis of LBP and low back diseases estimated that the annual total health care expenditure in Denmark amounted to approximately 16.8 billion DKK (3 billion USD), based on data collected in the period from May 2005 to March 2006 [34]. In another survey, the total costs for the Danish society were estimated to be 23.4 billion DKK (4 billion USD) in the period 2002-2003 [26].

Most cases of non-specific LBP appear to be relatively benign and are often described as 'self-limiting' for a few weeks, although recurrences are common [35]. Early identification of patients who are more likely to develop persistent disabling symptoms would help guide decisions regarding follow-up and management. Chou and Shekelle investigated the usefulness of individual risk factors for identifying these patients [36]. They concluded that the most helpful components for predicting persistent disabling LBP were maladaptive pain, coping behaviours, nonorganic signs, functional impairment, general health status, and presence of psychiatric comorbidities. Steenstra et al. systematically reviewed the literature dealing with prognostic factors for duration of sick leave in patients sick-listed with acute LBP and concluded that specific LBP, higher disability levels, older age, female gender, more social dysfunction and more social isolation, heavier work, and receiving higher financial compensation were identified as predictors of a longer duration of sick leave [37]. Exposure to spinal load is frequently discussed as a potential risk factor for LBP. It seems logical from a biomechanical point of view that high physical loads or working with one's trunk in a bent and/or twisted manner are associated with LBP, however the literature dealing with these relationships report conflicting conclusions [38-40]. Further conflicting evidence has also been reported about the potential association between chronic LBP and 'awkward postures' [41]. On the other hand, several reviews have consistently concluded that sitting, standing and walking are not risk factors for the onset of LBP [38;39;42;43]. A recent review explored the long-term associations between physical load and chronic LBP and found that among all the physical load variables, associations with chronic LBP were found only for awkward postures [44]. In another recent systematic review, Ramond et al. examined psychosocial risk factors for chronic back pain and concluded that a few psychosocial risk factors have been demonstrated to exist i.e. depression, psychological distress, passive coping strategies and fear-avoidance beliefs that have sometimes been found to be independently linked with poor outcome [20]. Research findings for other psychosocial risk factors are inconclusive, as in contrast to other studies, Hartvigsen et al. reported moderate evidence for no association between work-related social support or stress at work and outcomes [45]. Therefore, overall, it seems unclear if any robust risk factors exist for either LBP or chronic LBP.

Management of low back pain

“Too often, the choice of treatment reflects the skills of the professional rather than the needs of the patient. To put it simply, what treatment you receive depends more on who you go to see than on what is wrong with your back.” [21]

Clinicians and patients can choose from amongst numerous assessment and management options. Several national and international clinical guidelines for the management of different stages of LBP (acute, sub-acute or chronic) have been published in recent decades in an attempt to summarize the best available evidence and inform the best possible decisions in the clinic [2;11;25;26;46-48].

In general, the latest guidelines emphasize the importance of history-taking and physical examination for the process of diagnostic triage into classifications such as nonspecific LBP, radiculopathy or specific LBP (i.e. red flags, patients likely to have serious pathologies) and the assessment of psychosocial/prognostic/risk factors (yellow flags) [8;10;46;48;49]. Clinician/therapist-administered treatment for LBP can be categorized into education, medication, exercises, manual treatment and surgical procedures and should take into account the patient's needs and preferences [11]. When suspicion of serious disease (red flags) arises during examination, referral to a medical specialist is generally recommended and some guidelines also recommend referral in cases of no improvement following conservative treatment or aggravation of symptoms [8]. However if we look at the LBP where we cannot find a specific cause, we need to ask whether the LBP is truly “non-specific” or whether there may be further subtypes. Most care providers agree that chronic LBP is largely a mechanical or movement-related problem that may be influenced by social and psychological factors, but beyond this consensus, there is little agreement on etiology [50;51]. Likewise, the treatment approaches taken by different care providers show a great deal of variation [5;7]. This calls for useful classification systems that could guide clinicians in specific treatment strategies for chronic “non-specific” LBP patients. Currently, many classification systems exist for chronic LBP and overall these classification systems cover different focus areas (prognostic, treatment, diagnosis or more descriptive) [50;52;53]. In general, there seems to be insufficient evidence to recommend any classification system for non-surgical treatment of chronic LBP [53]. Classification and subgrouping based on dynamic motion characteristics has also been proposed using different systems and underlying theoretical models [54-59]. The results reported in the literature seem to be a promising indication that movement-based models for classifying LBP patients could improve the management of their pain. However in general, the literature in this field of science is limited to relatively few studies on smaller samples of people with LBP [54;55;57-60] highlighting that more research is required to further develop and validate these classification systems.

Manual therapy

There is a lack of precision surrounding manual therapy terminology in clinical research and it appears to depend mainly on the tradition and scope of the professional group utilizing it [61;62]. By definition however, ‘manual’ refers to the hands and the concept implies an application of physical treatments to evaluate, treat, and improve the status of neuromusculoskeletal conditions. Manual therapy may include several different treatment modalities used by various professional groups in the management of LBP [63;64]. Three commonly used manual therapy techniques are manipulative therapy, mobilization and massage, all based upon different application rationales and theories.

SMT or manipulation is defined in many different ways [61]. Shekelle defined spinal manipulation as ‘*a form of manual therapy that involves movement of a joint past its usual end range of motion but not past its anatomic range of motion, an area which is termed the “paraphysiologic zone”*’ [65]. Several treatment concepts have been developed such as the Diversified and Gonstead techniques, which use a variety of adjustment methods e.g. long lever low velocity (non-specific spinal) adjustments and short lever high velocity (specific spinal) adjustments. There are several pathophysiologic theories about underlying mechanisms presumed to be affected by manipulative therapy; some focusing on neurophysiological structures (changes in the neuromuscular system) and others on more biomechanically informed theories (e.g. trapped intra-articular synovial folds or meniscoids) [11;61;65-68].

Mobilization may be defined as the application of manual force to the spinal joints within the passive range of joint motion that does not involve a thrust [69]. Both manipulation and mobilization attempt to move the articular surfaces through passive forces. Whereas joint manipulation aims at obtaining a joint movement that may produce a cavitation (audible crack), mobilization passively moves the joint through its active and passive ranges of motion. To obtain a cavitation which is often viewed as indicating success, a force perpendicular to the articular surfaces must be generated [70]. For lumbar spinal manipulation, the forces have been estimated to be about 400N [71]. Other research, however, indicates that the distinction between manipulation and mobilization is probably not clear and cavitation may not be necessary to exert a clinical effect [72;73]. Clinical experience indicates that several manual therapies can have an immediate effect on pain. The literature offers many possible mechanisms and combinations of mechanisms to explain the pain-reducing effect of SMT and mobilization. However, the literature is conflicting and difficult to grasp because it consists of discussions, hypotheses and a mixture of studies employing different designs, methods and outcomes [74]. Overall, SMT has been hypothesised as having a direct effect on pain from three possible levels [74] i.e. local (mix of different mechanisms e.g. a decrease in the sensitivity of muscle spindles and/or the various segmental sites of a reflex pathway [75]), regional (effect at the spinal cord level i.e. where neural input is affected, subsequently altering

central processing and affecting reflex somatomotor or somatovisceral output [76;77]) or central (i.e. limiting the development of central sensitization of pain [77;78]). A recent review by Millan et al. indicates that a hypoalgesic effect of SMT is achievable and current evidence seems predominantly to support the theories based on local and regional effect levels although many questions still remain such as duration, magnitude of effects, precise mechanisms involved and if it is clinically significant [74]. Massage may be defined as soft tissue manipulation using hands or a mechanical device on any body part [79]. This technique depends on slower movements, pressure and stretch applied on the skin surface. The underlying biological and physiological effects of massage and their relationship with outcomes are mostly theoretical; however, pain relief may result from several mechanisms including stimulation of the autonomic nervous system, an associated release of endorphins or pain inhibition mechanisms [80-82].

Several guideline recommendations have been developed in the last decade regarding the benefits of manual therapies for the care of LBP [2;11;48;49;83]. In 2010, Bronfort et al. published a comprehensive summary report of the scientific evidence regarding the effectiveness of manual treatment for a variety of conditions including LBP [84]. The conclusions of the report were based on systematic reviews of randomized control trials (RCTs), clinical guidelines or technology assessment reports and all RCTs not yet included in those reviews and reports. It was concluded that for LBP, manipulation or mobilization therapies were effective treatment options for sub-acute and chronic LBP in adults (high quality evidence) and acute LBP in adults (moderate quality evidence). Adding spinal mobilization to medical care does not improve outcomes for acute LBP in adults (moderate quality evidence). Massage is an effective treatment for sub-acute and chronic LBP in adults (moderate quality evidence). Furlan et al. concluded that massage is beneficial for patients with sub-acute and chronic non-specific LBP in terms of improving symptoms and function and that the effects of massage are improved if combined with exercise and education [79]. In a recent systematic review and meta-analysis by Furlan et al., it was concluded that complementary and alternative medicine (CAM) treatments (acupuncture, massage, spinal manipulation, and mobilization) were significantly more effective than physical therapy, placebo, no treatment, or usual care in reducing pain immediately or shortly after treatment. However none of the CAM treatments was systematically shown to be superior to any other and no significant reduction in disability compared to sham was found [85].

Several studies have assessed the effectiveness of different intervention modalities. SMT has been recognized in several systematic reviews and clinical guidelines as a treatment option for patients with chronic LBP [11;84;86-88]. The overall message is that SMT can bring about pain relief and improve functional status and is therefore

recommended, even though the effect may be small. Any clinically relevant difference in the effectiveness of SMT compared to other treatment alternatives such as exercise therapy seems to be questionable [87;89].

In a recent systematic review by Slater et al., the effectiveness of subgroup-specific manual therapy for mechanical LBP was assessed (excluding trials where LBP was due to serious or non-mechanical pathologies). Significant treatment effects were found to favor subgroup-specific treatments [90]. Another review by Kent et al. also concluded that the available evidence for treatment (exercise therapy and manual therapy) targeted to subgroups of patients with non-specific LBP for improving patient outcomes was weak and had to be interpreted with caution. However adequately powered controlled trials using designs capable of providing robust information on treatment effect modification are uncommon [6].

Self-care interventions and education

Van den Borne has defined patient education as *'a systematic experience in which a combination of methods is generally used, such as the provision of information and advice and behaviour modification techniques, which influence the way the patient experiences his illness and/or his knowledge and health behaviour, aimed at improving or maintaining or learning to cope with a condition, usually a chronic one'* [91].

The general consensus is that using education to manage all stages of nonspecific LBP and keeping the patient active is beneficial [10;46]. Clinical guidelines recommend that non-specific acute LBP patients should receive self-care options, remain active and return to work as soon as feasible. For the patients who do not improve with self-care, therapists are advised to consider additional therapies/treatments such as exercise and spinal manipulation [8;83].

Although self-care is recommended as a first option, it has been indicated that in practice, patients with LBP spend substantial amounts of money on therapy soon after the initial LBP diagnosis [92]. Self-care and education is likely to be less costly and time-consuming than other therapies and might therefore be an attractive and efficient intervention modality not only for the patient with LBP but also for the society in which they live.

A systematic review has shown strong evidence that individual education for LBP patients may be effective for acute and sub-acute LBP; however, for chronic LBP the evidence is less conclusive [93]. Less than 2.5 hours of individual oral educational sessions were no more effective than no intervention for sub-acute LBP. They concluded that more research is needed to confirm these results, and to find out which types of patient education

are the most effective [92]. People with chronic LBP have been found to be less likely to benefit from patient education than people with acute LBP [93].

Exercise therapy

Exercise has been described as *‘a form of physical activity that is planned, structured, repetitive, and purposeful with a main objective of improvement or maintenance of one or more components of physical fitness i.e. all exercise is physical activity, not all physical activity is exercise’*[94].

Exercise therapy encompasses a heterogeneous group of interventions and may include a variety of strengthening exercises (e.g. core strengthening [95]), coordination (e.g. Alexander technique [96;97]), stretching/flexibility/directional programs (e.g. McKenzie method[98]), home-based programs [1;99] and aerobic exercises designed to increase baseline physical activity levels [100].

Several reviews and guidelines indicate strong evidence supporting the effectiveness of exercise therapy in chronic LBP patients and moderate or unclear evidence for its ineffectiveness in acute LBP [8;10;101-103]. For acute LBP, generalized exercise therapy compared to other conservative treatments has not clearly demonstrated a treatment benefit [48;104]. One systematic review of exercise in patients with acute LBP found that exercise therapy was no more effective than no treatment or other conservative treatments which included NSAIDs/other analgesics, patient education programs, and/or advice to stay active [104]. It is commonly recommended in guidelines that patients who have chronic LBP perform physical, therapeutic, or recreational exercise, keeping in mind that no specific active technique or method is superior to another. However Hayden et al. reviewed the literature in order to identify particular exercise intervention characteristics that most effectively decrease pain and improve function in adults with non-specific chronic LBP [102]. They classified exercise therapy according to program design (individual or standardized), delivery type (with or without supervision), and dose (high or low). They concluded that individually designed and supervised programs of stretching and strengthening, encouraging adherence to achieve high dosage, seems to be the most effective [102].

The general health benefits from physical activity are considerable [105], however sedentary behavior is widely prevalent [106]. There are, therefore, many good reasons for motivating patients with sedentary lifestyles to become more active. However, for chronic LBP patients, a recent review by Griffin et al. concluded that there is no conclusive evidence that patients with chronic LBP are less active than healthy individuals [107]. This finding may be a little surprising given the fear-avoidance model [108]. Nevertheless, benefits from exercise have been documented into late adolescence [109] and Hartvigsen et al. concluded that strenuous physical activity at least once a week is protective for incident LBP in elderly people [110]. For people with acute, sub-

acute or chronic LBP, there is no evidence that exercise increases the risk of additional back problems or work disability. To the contrary, it might slightly reduce the risk of future back injuries [111]. Exercise can be used with the goal of improving impaired capabilities such as flexibility, back strength, cardiovascular endurance or reducing chronic pain symptoms and behavioral disabilities [111].

Physiology of pain

Pain can be defined as “*an unpleasant sensory and emotional experience associated with actual or potential tissue damage, or described in terms of such damage*” [112]. Knowledge of pain physiology is essential in order to understand and further develop pain management strategies. The pathway of pain goes from nociceptors (free nerve endings reacting to biological, mechanical, electrical, thermal or chemical stimuli) located in all tissues of the body (except the brain) through the peripheral nervous system (myelinated A-fibers and unmyelinated C-fibers) to the central nervous system and up to the brain (from the thalamus to limbic system and to the cerebral cortex)[113]. Within the spinal horn of the spinal cord, the pain fibers synapse with spinal neurons via synaptic transmission [114]. Acute pain serves a purpose by providing a warning that an injury is occurring; however pain is an extremely complex interaction of both physical events and affective and cognitive traits of the person who experiences it. For instance, research has indicated that patients’ expectations regarding the analgesic effect of a drug can change the drug’s analgesic effect i.e. anti-analgesic expectations may block the action of the drug in the spine [115].

Several pain control theories have been proposed, however the best known might be the “gate control theory of pain” by Melzack and Wall in 1965[116]. Gate control theory is one explanation for the modulation of pain in the spinal cord. According to this theory, a balance between myelinated afferent nerve fibers, that are not directly related to the nociceptive transmission, and the nociceptive afferents modulate the output of the central transmission of the noxious input that determines the perception i.e. the intensity of the pain[117]. In plain language, an “open” gate permits the flow of nerve impulses from the peripheral nervous system (PNS) through the central nervous system (CNS) to the brain, and as a result, the brain is able to perceive pain. A “closed” gate does not permit this flow. Although the control gate theory is widely accepted, it has unanswered questions regarding e.g. chronic pain issues, gender-based differences and the effects of previous pain experiences. Chronic pain is more complex than acute pain making it very difficult to understand and indeed challenging to manage [113]. Therefore Melzack and Wall came up with an improved theory in 1999, which was called the “neuromatrix theory” [118]. This theory included the hypothesis that each person has a unique matrix of neurons that is affected by all facets of biopsychosocial traits i.e. pain is considered a complex multidimensional interaction and not simply a relationship between a damaged tissue and pain.

Painful experiences may imprint themselves on the nervous system. Prolonged or strong activity of neurons in the dorsal horns by sustained or repeated noxious stimulation may lead to increased neuronal responsiveness or central sensitization which may cause exaggerated perception (hyperalgesia and allodynia)[114;119;120]. One of the mechanisms behind central sensitization has been described as the wind-up phenomenon. It is a progressive increase of the duration and amplitude of action potentials in the neurons of the dorsal horn (causing increased pain experience) after a repetitive and constant stimulation of C fibers [114;121]. Overall pain and chronic pain perception may be modulated (exaggerated or diminished) by messages according to different biological or pharmacological substances affecting different sites in the PNS or CNS, psychosocial factors originated at the brain and through motor activity (movement) [113].

Mechanisms linking movements and pain

Potential influences facilitating the benefits from physical activity may be related to biological, psychological and social factors; however in this section, mainly the biological mechanisms will be appraised. It is known that pain negatively impacts physical behavior and it is also well documented that the ‘right’ dose of physical activity positively impacts pain [101;103;111;122]. However the knowledge of the biological link between movement and pain is limited.

“In pain populations there may be a threshold at which the benefits of exercise and fatigue become detrimental.” [123]

Most people have woken up the day after unfamiliar hard physical work and felt pain and soreness in several different body locations (e.g. joints and muscles). When we try to get out of the bed and start to move around, we can feel that our movements and movement patterns have changed partly in response to the pain. In this case, certain unaccustomed activities such as high-force eccentric exercise might have provoked muscle strains (delayed onset muscle soreness) and pain. Even though movement or physical activity can induce pain if inappropriate intensity or type is used; physical activity is frequently used in the clinical setting as part of pain management. A recent review by Cote and Hoeger indicates that exercise prescription in patients with musculoskeletal pain is difficult because of the limited amount of high-quality research, differences between pain conditions, and diversity within the same pain condition [123]. For chronic LBP patients, a range of different types of physical activities seem to be relatively equally beneficial e.g. strengthening, aerobic, stabilizing exercise or coordination [124-128]. One way to interpret this might be that for chronic LBP, the specific type of physical activity might not be as important as increasing the general activity. However, the optimal intensity and duration of physical activity may be linked to the pain condition being treated although this is not yet clear [123]. In addition, if hypoanalgesic mechanisms work from different sites that can be affected by

different kinds of physical activity, this could be used in deciding or combining treatment therapies to get the best effect.

The mechanisms through which exercise may reduce pain are currently not well established. This, in turn, makes it difficult for clinicians to direct specific treatment strategies in order to have the best possible pain-reducing effect for the patient. It is difficult to establish a good overview of how movement interventions could reduce pain because of the multidimensionality and complexity of both the pain conditions and the interventions e.g. what type of activity/exercise, at what dose, for which patient? The effect of a given movement intervention may be dependent on the type of pain condition. Nevertheless enhanced knowledge of the underlying exercise-induced hypoalgesia (EIH) mechanisms may help clinicians and researchers to specify the movement intervention in order to optimize the effect. EIH may come through different biological pathways which can be divided into opioid and non-opioid mechanisms [123]. Research has shown that activation of the opioid system (increase the plasma level β -endorphin level; which is known to be anti-nociceptive) is possible through physical activity which may also explain the decrease in pain [129;130]. However, there seems to be a limited number of studies especially on chronic pain patients and interpretations should be carried out carefully also due to contradictory results [123]. In addition, most studies show no correlation between the β -endorphin release and the EIH response which leads to questions concerning how direct the link is [123;129;131]. Any knowledge of potential biological non-opioid mechanisms between physical activity and EIH are severely lacking [123]. However in the few studies available, some indications and hypotheses have been generated. One neurological mechanism was proposed by Rainville et al. who hypothesized that exercise may reduce pain through desensitization (neurological or physiological) of the pain-producing tissue, by way of the repeated application of force or stress to that tissue [111;132]. Another specific biological or mechanical factor may be that moderate dynamic loading of the spine during exercise therapy may facilitate diffusion of nutrition and an anabolic effect on the intervertebral disc matrix [133;134] which in turn also may influence pain. One neurophysiological explanation may be that activation of the motor pathway, through physical activity, directly modulates the pain pathway. This may be the result of descending inhibition or that physical activity influences the perception of somato-sensory stimuli including pain [123;135;136]. Physical activity has the potential to influence all aspects of the biopsychosocial model, which increases the complexity in the process of identifying responsible mechanisms for EIH.

Biomechanical measurements of the lumbar region

For many years, researchers and clinicians have sought to measure back problems objectively, primarily to attempt to determine the origin of pain, and subsequently, to measure deteriorations or improvements in the

condition, e.g. if given types of treatment evoke biologically or biomechanically measurable changes that are related to changes sensed by the patient.

Anatomical diagnosis of LBP conditions on the basis of a presumed injured or painful structure using advanced imaging techniques has been shown to be possible only in a minority of patients [59;137;138]. Many asymptomatic individuals have evidence of pathology, and symptomatic patients may not necessarily have identifiable pathoanatomical sources. The inadequacies of pathoanatomical diagnosis of LBP led to the development of the Quebec Task Force Classification system which is based on patient symptoms and neurological signs [139]. However, it is questionable as to how much this system has improved our understanding of the mechanisms of LBP or informed our choice of treatment [140].

Functional capacity assessments addressing strength and endurance of trunk musculature are quantitative attempts to measure function in LBP patients [141]. However, it may not be appropriate to measure ‘extreme’ capacity of e.g. back strength in highly disabled patients, as it often goes far beyond the ‘normal’ trunk function needed for daily activity [60].

The development of techniques to measure trunk motion characteristics in unloaded free dynamic activities is an attempt to remedy the existing deficiencies of LBP impairment quantification [142-145]. The theoretical reason for looking at spinal function has often been based on the assumption that some ‘mechanical abnormality’ such as degeneration or prolapsed intervertebral disc, spondylolisthesis or other pathology led to an abnormal function [144;146]. Another argument for assessing motion has been to identify patients who are exaggerating their pain or making excuses not to work [147]. Bishop et al. stated that “*kinematic assessments are attractive because a kinematic abnormality may reflect the underlying disease*” [54]. McGregor et al. investigated how classification based on diagnostic groups (i.e. disc prolapse, degenerative disc disease, spondylolisthesis, stenosis and non-specific LBP) related to lumbar spinal motion characteristics [148]. The study showed that there were significant differences in the motion characteristics between the normal and diagnostic groups and, in addition, that some of the variability between the diagnostic groups could be attributed to the different pathological processes. It was proposed that motion analysis could be used as a surgical indicator or as an outcome measure in surgical evaluation.

Some technologies have been designed to investigate the motion of each vertebra (segmental motion) i.e. digitized videofluoroscopic images [149] or by invasive methods [150;151], and others are designed to measure the spine at a regional level e.g. the entire lumbar spine [142;152]. The segmental and invasive motion assessments may be considered clinically or scientifically valuable, but have been considered to be beyond the

scope of this thesis, which focuses on non-invasive equipment that is attached to the skin surface to measure regional lumbar motion.

Functional evaluations of patients suffering from low back disorders are a major challenge in daily clinical practice. In the clinic, regional motion examination using simple low-tech measurements such as Inclinator ROM, finger-to-floor distance or Schober's Index are common. Back and forward bending are important components of many functional activities, and are routinely assessed in the clinical evaluation of LBP. However, motions in other directions may also be important, e.g. investigation of asymmetry in lateral flexion or rotation [153].

New technologies (non-invasive, real-time, three-dimensional (3D), regional instruments) have made it possible to derive and quantify more sophisticated motion parameters such as motion velocity, acceleration, symmetry or motion area [54;57;58;142]. Conceptually, these measurement instruments are intended to measure the same construct while using different underlying technologies e.g. electromagnetic [154], potentiometric [155;156], gyro-metric [157], opto-electronic [158] as well as ultrasound pulse [159] measuring devices. These non-invasive instruments have no known side effects, and compared to roentgenographic analysis, these instruments carry no risk of exposure to radiation. It is assumed that a large proportion of LBP is caused or significantly influenced by biomechanical factors [13;54;59;160] making a functional assessment an obvious choice in order to e.g. differentiate between subtypes and or evaluate progress over time[143].

Lumbar kinematics and patient-rated outcomes

“To find an objective and quantifiable method to measure disability in patients with LBP, clinicians and researchers have turned to spine kinematics.” [161]

Kinematics is a subdivision of biomechanics that studies movements independently of forces. The most commonly used kinematic element is ROM; however velocity and acceleration are other examples.

The way in which a person uses his/her back may determine the presence or absence of pain [13]. One potential attraction of kinematic assessments is the notion that they might display abnormalities reflective of an underlying disease [54;148]. For instance, people suffering from back pain may avoid certain postures that cause pain, or similarly, muscle activation patterns may be altered because of pain.

Several measurements used in clinical settings to evaluate LBP in patients rely on spine kinematics. Some commonly used measurements, as mentioned in the section above, are Inclinator ROM, finger-to-floor distance or Schober's Index. These tests quantify end range of trunk movement by simple means; however the

associations between these low-tech ROM measures and patient-rated measures such as pain and disability questionnaires have often been reported to be low [162-165].

The development of new technologies has made it possible to derive and quantify more sophisticated motion parameters such as motion velocity, acceleration, symmetry or motion area [54;57;58]. These new motion parameters may show more promise.

In a literature review, Lehman stated:

“Assessing the higher order kinematics of patients before and during rehabilitation may provide a better means of documenting patients’ progress compared with simple pain measures. It may also be an excellent means of elucidating the mechanisms of individual treatments.”[142]

In other words, if spine kinematics is related to back pain, then patients suffering from back pain should show a change in kinematic variables over time if their condition is improving.

As recommended by McGregor and Hughes *“Further work in this area with large study populations and less subject drop-out rates is required to investigate in greater detail the relationship between pain and function, particularly with respect to different spinal pathologies” [145].*

Reliability and validity

‘An essential requirement of all measurements in clinical practice and research is that they are reliable.’ [166]

In articles and textbooks on reliability, a variety of synonyms are used. Commonly used synonyms include reproducibility, repeatability, precision, accuracy, concordance and agreement. In addition, the definition of reliability varies considerably. The COSMIN group defines reliability as *‘the degree to which the measurement is free from measurement error’* [167].

The COSMIN group divided the reliability domain into three measurement properties: (1) internal consistency, (2) reliability and (3) measurement error. Therefore the term ‘reliability’ is used twice, firstly as the term for the domain and secondly as the term for the measurement property [167].

Internal consistency is irrelevant in relation to motion measures such as ROM and has not been used in the thesis. The measurement properties of reliability and measurement error have been defined in the following way by the COSMIN group:

- Reliability: The proportion of the total variance in the measurements that is due to ‘true’ differences among patients.
- Measurement error: The systematic and random error of a patient’s score that is not attributed to true changes in the construct [167].

In terms of reliability measures, the Intraclass Correlation Coefficient (ICC) provides a unitless index based on the ratio of the within-to-between subject test-retest differences and ranges from 0.00 to 1.00, with values closer to 1.00 representing stronger reliability. Reliability may be defined in a range of ways; however, they all refer to the basic formula from classical test theory:

$$Reliability = \frac{\sigma_p^2}{\sigma_p^2 + \sigma_e^2}$$

where σ is the standard deviation (SD) and σ^2 is the variance, ‘p’ denominates subjects (σ_p^2 between-person variance) and ‘e’ relates to measurement error (σ_e^2 within-person variance). Therefore the reliability is the between-person variance divided by the between-person variance plus the within-person variance (measurement error).

The coefficient in itself gives no indication of the magnitude of the measurement error but rather an indication of the extent to which a measurement instrument can differentiate among individuals, despite measurement error. There are several factors that could influence reliability and measurement error. These factors may manifest as random variance or systematic error. Some ICC formulae do not include the variance for the systematic difference (systematic bias) as a source of error, and thereby, in general, generate inflated values. Therefore, a comparison between reliability values from studies with similar study designs but different statistical analyses is problematic.

There are no standard criteria for acceptable reliability and therefore ICC values reported in the literature may be difficult to interpret. As a guideline, Portney and Watkins suggest that values below 0.50 represent poor reliability, coefficients from 0.50 to 0.75 suggest moderate reliability, and values above 0.75 are indicative of good or high reliability [168]. Aaronson et al. recommend that minimal standards for reproducibility coefficients be 0.70 for group comparisons and 0.90–0.95 for individual measurements over time [169]. The criteria used by a researcher or clinician must be based on how the results of the reliability test will be applied. In research, we may be able to tolerate lower reliability if measures are used for evaluation of group mean changes than if the measures are used for diagnostic purposes or decision-making.

The error variance is an index of how far apart the repeated measures are. If you take the square root of the error variance, you get the standard error of the measurement (SEM) which is a parameter of measurement error [166].

In 1986, Bland and Altman proposed another method for assessing measurement error [170]. This method, called the limits of agreement (LOA), has become very popular, with the paper this year exceeding 20,000 citations. The authors describe a relatively simple statistical method and graph for using paired data to assess the differences between measurements obtained by two different measurement systems or to compare two measurements obtained by the same method. In the plot, the difference between the measurements is displayed against the mean of the same measurements with 95% LOA calculated as $\text{mean difference} \pm 1.96 * \text{SD}_{\text{difference}}$. In this way, the 95% LOA provide an interval within which 95% of the difference between the two measurements is expected to lie [171]. To evaluate changes over time in an individual, the change must exceed the inherent variability of the repeated measurements, which can be determined using the LOA.

Measurement validity concerns the extent to which an instrument measures what it is intended to measure. For instance, a goniometer may be considered a valid (face validity) instrument for measuring lumbar ROM because we can assess joint range from angular measurements. Reliability sets the limits of validity, although high reliability does not automatically imply high validity.

“Validity implies that a measurement is relatively free from error; that is, a valid test is also reliable.”

[168]

Validity can be defined as *‘the degree to which an instrument truly measures the construct(s) it purports to measure’* [166]. Generally three types of validity can be distinguished (with numerous subtypes). Content validity can be defined as *‘the degree to which the content of a measurement instrument is an adequate reflection of the construct to be measured’* [166]. Criterion validity can be defined as *‘the degree to which the scores of a measurement instrument are an adequate reflection of a gold standard’* and construct validity as *‘the degree to which the scores of a measurement instrument are consistent with hypotheses, e.g. with regard to internal relationships, relationships with scores of other instruments or differences between relevant groups’* [166].

As an example, we could compare spinal ROM measures on the same subjects from two different instruments e.g. goniometer vs. x-ray. If we consider the x-ray measurement as the gold standard (criterion validity) we could test if the outcomes of the goniometer can be used as a substitute measure for the x-ray. In this way, validity addresses what we are able to do with test results. Research has indicated that different measurement systems used in the assessment of lumbar spinal movement might yield non-comparable values for the same spinal movement due to differences in either the manner in which the device is attached to the participant, or the accuracy with which the device records the movements in the given plane [172;173]. However, when assessing longitudinal changes in an individual’s mobility using the same instrument e.g. in monitoring progress during rehabilitation, it is of primary importance to ensure that the device itself yields precise measurements and that reliable outcomes can be obtained using the same instrument. The concept of instrument accuracy and precision is often used when a researcher describes or examines an instrument measurement error relative to itself (precision) or relative to another gold standard instrument (accuracy) without the human error from the interaction. Overall, the equipment used for kinematic assessment can be considered relatively precise and accurate. The biggest amount of measurement error seems to be related to human interaction.

Aims and hypothesis of the thesis

This PhD project investigates regional lumbar motion measurements and their application in the assessment of chronic LBP patients. Three of the four manuscripts included in this PhD thesis are based on secondary analyses of objective motion measurements recorded at the Wolfe-Harris Center for Clinical Studies at the Northwestern Health Sciences University in Bloomington, MN, USA in conjunction with a randomized clinical trial [1]. Therefore, these studies are related to secondary aims established prior to the principal data collection. The overall aims of this thesis were to (1) obtain further understanding of the reliability of measurements for regional lumbar motion, and (2) examine motion changes over time and the association between change in objectively measured regional lumbar motion and change in patient-rated pain and back-related function.

Specific objectives and hypothesis:

1. To systematically and critically review the literature on regional lumbar motion measurement systems and evaluate the quality of the reporting on the reliability and/or measurement error of these systems (Manuscript I).
2. To evaluate the reliability and measurement error of regional lumbar motion assessed in chronic LBP patients and to quantify underlying sources leading to measurement error between repeated measurements (Manuscript II). Based on the reliability review (Manuscript I[174]), we hypothesized that reliability for all motion parameters would have an $ICC_{(1,1)}$ in the range of (0.6-0.8) in participants with a stable pain level.
3. To quantify the effect of 12 weeks of SMT, SET or HEA on regional lumbar motion in chronic LBP patients (Manuscript III). We hypothesized that the groups receiving either SMT or SET would change significantly in all motion parameters (specifically, the Jerk Index was hypothesized to decrease and all other motion parameters to increase) over a 12-week period, whereas the minimal intervention group (HEA) would not change in motion parameters.
4. To compare change in objectively measured regional lumbar motion to change in pain and back-related function over a 12-week period, for the whole study sample (Manuscript IV). We hypothesized that if patients reported a clinically relevant improvement in patient-rated pain and function (i.e. >30%), this would correspond to greater change in objectively measured motion parameters compared to patients who did not report a clinically relevant improvement in self-reported pain and back-related function. Further to this an exploratory analysis was conducted to investigate whether this relationship was similar in subgroups: (i) based on pain distribution, (back pain only versus back and leg pain), and (ii) receiving different treatments (SMT, SET or HEA over a 12-week period).

Method

Summary of methods and materials

This thesis consists of four manuscripts. The first is a literature review and the remaining three examine motion data recorded in a single cohort of chronic LBP patients. Manuscripts II, III and IV are secondary analyses of a subset of participants from a RCT conducted by Bronfort et al. [1]. Bronfort and co-workers at the Wolfe Harris Center for Clinical Studies at Northwestern Health Sciences University, Minneapolis, MN, USA, collected the motion data over a period of 3 years. The institutional review boards of the Northwestern Health Sciences University, the Minneapolis Medical Research Foundation, and the University of Minnesota approved the study (RCT) and written informed consent was obtained from all study participants. Critical methodological considerations will be evaluated in the discussion section at an overall level and in more detail in each of the Manuscripts I to IV.

The systematic review

A search strategy was developed and used to search four databases (Pubmed, CINAHL, Embase and Mantis) for relevant articles published prior to May 2011. The search strategy was developed in collaboration with a health science research librarian and the following search terms were used: (lumbar OR lumbosacral OR lumbal) AND (spinal OR spine OR vertebrae OR vertebral) AND (motion OR biomechanical OR biomechanics OR kinematic OR kinematics) AND (analysis OR analyzing OR analyzer OR measurements OR measuring OR measure OR measures OR assess OR assessed OR assessment) AND (velocity OR ROM OR "range of motion" OR acceleration OR accelerometer) AND (pattern OR patterns OR coupled OR flexion OR extension OR rotation OR rotations OR lateral). In addition, reference lists were searched for further studies.

Inclusion and exclusion criteria

The review included original articles dealing with test-retest reliability/reproducibility of a 3D computerized regional lumbar motion analysis system in human subjects. Studies had to be published in peer-reviewed journals as full articles and written in English, Danish, Swedish or Norwegian. The original studies had to report on the collection of non-invasive 3D data, recorded electronically in real-time under standardized conditions.

Data collection and management

All retrieved titles and abstracts were inspected by the PhD candidate, who excluded all articles that did not meet the inclusion criteria. The two authors who entered data into the checklists independently of each other reviewed all included articles. The checklists were compared and inconsistencies were resolved by consensus.

No fixed set of generally accepted quality criteria were found that suited this type of literature review and therefore a checklist was designed that included descriptive items and items for quality assessment. The checklist was developed drawing on information from previous reviews [175;176] and guidelines for reporting reliability and agreement studies (GRRAS) [169;177-179], the standards for reporting of diagnostic accuracy (STARD) [180;181] and the quality assessment of diagnostic accuracy studies (QUADAS) [182-184]. The checklist was designed to guide the systematic extraction of information in a standardized way. The descriptive items were then used as a foundation for the quality assessment of the included studies. The checklists were divided into four domains: study population, testing circumstances, equipment and data analysis/presentation. The quality assessment was designed to summarize each domain in one expression of completeness of reporting based on all of the descriptive items. Categories for each domain were 'yes', 'partly' and 'no'. Thus, the best quality assessment evaluation for a study would be 'yes' in all domains.

The criteria for the scores were: 'Yes' if the information was found to be complete or very close to complete, 'Partly' if some but not all information was missing and the reporting in general was considered insufficient, and 'No' when there were major deficiencies or no information was reported.

For each of the four domains, the specific quality criteria were:

- The study sample represents a well-defined population, and description of participants is sufficient to replicate the study group
- The description of testing procedure and circumstances is sufficient for others to replicate the procedures
- The description of the equipment is sufficient for others to assess the technology
- Data presentation is reported in sufficient detail for others to assess the results and the statistical methods.

Analysis and synthesis

All extracted descriptive information was synthesised and presented in tables. The information provided by the descriptive extraction was used in the judgement of quality of each domain and entered into the tables as yes, partly or no. Finally, the quality of individual studies and the overall evidence was summarized in tables and figures for each group of instruments.

The cohort studies

Experimental protocol and examiners

Regional lumbar motion recordings were measured during two baseline visits (separated by 7-14 days) and one follow-up visit after 12 weeks of intervention and self-reported outcome measures data were collected at baseline and 4, 12, 26 and 52 weeks post-randomization. Data collection was performed by chiropractic and medical investigators who had worked together for several years. The investigators worked closely throughout the course of this research project. This included monthly meetings discussing protocol and policy decisions, examination of patients together during the baseline clinical evaluation of patients, referral of potential participants from medical investigators' clinical practices, review and referral of problematic cases.

Nine trained and certified research clinicians performed the objective evaluation and outcome assessment while blinded to the clinical information. For logistic reasons, it was not possible for the same examiner to conduct the examination of the same patient on each occasion.

On the first visit, participants' anthropometrical data (height, weight) were obtained and all subjects completed a self-administered questionnaire seeking information on their health history and demographics. Subsequently, chiropractic and medical clinicians reviewed the health history and performed a physical examination including a complete neurological examination, orthopedic tests, and manual static and motion palpation of the lumbar spine and lower extremities. Participants who qualified and agreed to participate were then scheduled for a second baseline visit in the clinic where the test procedures were repeated.

Study sample

Three hundred and one patients were included in the original RCT [1]. Due to technical problems with the equipment at baseline or follow-up and because of dropouts, different numbers of participants were included in the studies presented in this thesis (Figure 2). For inclusion in the reliability study (Manuscript II), complete motion data from the two baseline trials were required, which resulted in the inclusion of 220 patients. For inclusion in the intervention studies (Manuscripts III and IV), participants must have completed the second baseline assessment, provided follow-up regional lumbar motion data and have participated in the RCT. This resulted in the inclusion of 199 patients.

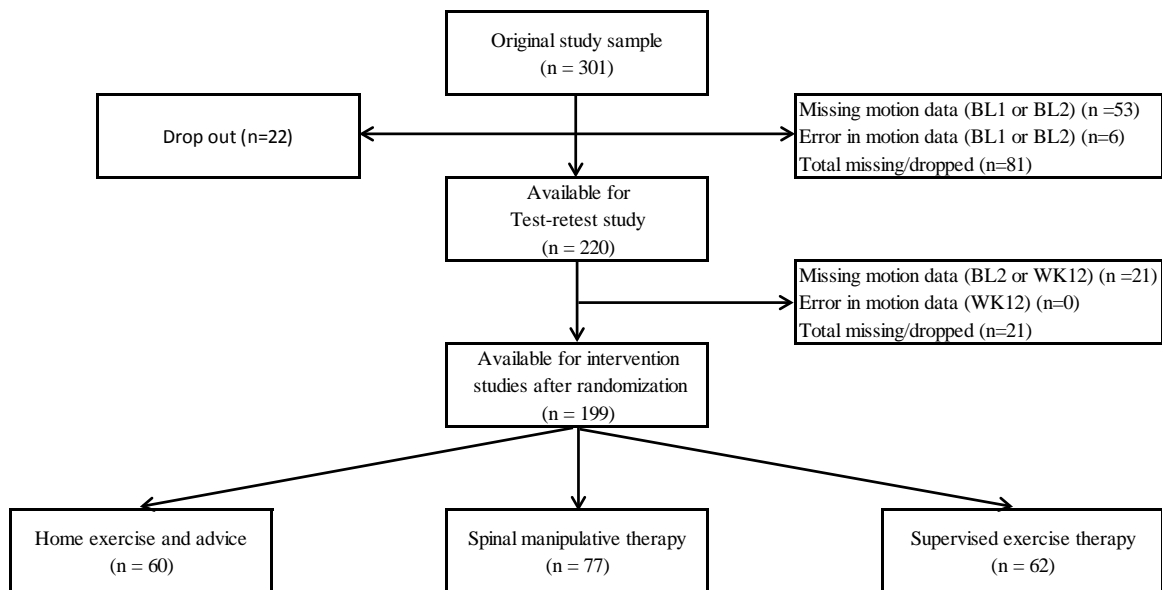


Figure 2: Flow diagram illustrating the participants available for test-retest analysis and intervention studies.

Inclusion and exclusion criteria

The RCT had the following inclusion criteria: 18 to 65 years of age with a primary complaint of mechanical LBP of at least 6 weeks' duration, with or without radiating pain to the lower extremity. Mechanical LBP was defined as pain that had no specific identifiable etiology but that could be reproduced by back movements or provocation tests.

The RCT exclusion criteria were: people with previous lumbar spine fusion surgery, progressive neurological deficits, aortic or peripheral vascular disease, pain scores of less than 3 (0–10 scale), involvement in pending or current litigation, or ongoing treatment for back pain by other health care providers.

These criteria were implemented in order to achieve a homogeneous and stable cohort with consistent symptoms and severity of LBP, where the level of pain intensity made it possible to measure changes over time and where other health conditions were unlikely to influence the outcomes over the one-year follow-up period. Participants were recruited principally through local newspaper advertisements, community posters, and postcard mailings. Initial screenings were conducted by telephone.

Randomization and blinding

In the original study, restricted randomization using a 1:1:1 allocation ratio was applied using four strata: patients with radiating symptoms, patients without radiating symptoms, LBP of 6 to 12 weeks' duration, and

LBP for more than 12 weeks. Prior to enrollment, the project statistician generated a randomization list using randomly mixed permuted blocks of different sizes. Examiners masked to treatment allocation performed the objective outcome assessment.

Interventions

Clinicians used standardized forms to document the events and procedures of each treatment visit, including patient-rated side-effects. A minimum of 80% attendance at the scheduled visits was required to be considered compliant with the treatment. The following intervention modalities were employed: SMT, SET and HEA.

Spinal manipulative therapy (SMT)

The number of treatments and the schedule of care were determined by one of the nine treating chiropractors. Treatment typically involved two encounters per week lasting 15-30 minutes that could include manual spinal manipulation, light soft tissue massage, with the assistance of a flexion/distraction table if required. Activity modification was prescribed as necessary. The individual clinicians determined the vertebral levels treated after using static and/or motion palpation. Specific spinal manipulation was performed as follows: patients were positioned on a treatment table in either the prone, supine or side-lying position. For each spinal manipulation, the chiropractor's contact hand would be placed over an osseous process, muscle or ligament and the vertebral or sacroiliac joint of interest would be taken to the end of its physiological ROM. The chiropractor would then apply a high velocity, low amplitude impulse to the joint, carrying it beyond the normal physiological ROM. Participants were discharged from care if the treating clinician felt that maximum clinical benefit had been obtained. The average number of spinal manipulative treatments was slightly more than 16.

Supervised exercise therapy (SET)

Supervised high dose exercise in small groups of patients (3 to 4) was provided (one-on-one supervision) by 15 exercise therapists trained in the study protocol. The main focus was dynamic trunk strengthening exercises (trunk extensions and leg extensions) and abdominal exercises using low-tech methods. The main goal of the program was to increase trunk muscle endurance and trunk stability [124]. In addition, a core strengthening program and static stretches (series of six) were implemented with a focus on the lumbar, gluteal, and hamstring musculature before and after strengthening. Each stretch was done once, with the patients instructed to hold each stretch for three deep breaths. Over the 12-week period, patients were asked to attend 20 one-hour sessions involving a high number of repetitions (two to three sets of 15–30 repetitions for each exercise) and a progressive increase in muscle load (achieved by altering the patient's center of gravity when possible). The patients were instructed to perform repetitions until they could no longer do so using proper form. The exercise

was classified according to type, program design, delivery, and dose as described by Hayden et al. [102]. The study protocol has, in part, been tested in a previous trial [185].

Home exercise and advice (HEA)

Eleven therapists who were trained in the study protocol provided counseling on self-care education. Two one-hour sessions were conducted on self-care measures and ergonomics associated with work and activities of daily living. Individualized sessions included advice and instruction on self-care measures, such as the use of ice and heat, ergonomic recommendations for home and work, and demonstration of good lifting techniques. Simple stretching and strengthening exercises, including lumbar extension, bridging, and abdominal crunches, were demonstrated and practiced with patient participation. Study participants were given a book and laminated cards describing these exercises and were encouraged to perform them at home on a daily basis. The patients were followed up in person 1 to 2 weeks later and then instructed to continue with the exercises on their own for the remainder of the intervention phase. The program was considered to be of low dose because of the simplicity of the exercises, time required to perform them (2–3 minutes per series), and low number of health care provider visits.

Data collection and variables of interest

Evaluation was conducted during two baseline assessments and 12 weeks after randomization. Lumbar motion data were collected by blinded examiners. Patient-rated pain was measured on an ordinal 11-box scale [186] where the patients were asked to rate their typical level of back pain over the previous week on a 0-10 scale, with 0 being 'no pain' and 10 being 'worst pain possible'. Besides this, no other pain measurements were assessed during the objective regional lumbar motion test procedure. Back-related function was measured on a Modified Roland Morris Disability Questionnaire (RMDQ) [187]. A range of other tests were performed i.e. isometric trunk flexion and extension strength and endurance, and patient-rated outcomes were collected including quality of life (36-Item Short Form Health Survey, version 2), frequency of pain (9-point ordinal scale), medication use for their LBP over the previous week, patient-perceived global improvement, depression (Center for Epidemiologic Studies Depression Scale - CESD), Fear-Avoidance Beliefs Questionnaire (FABQ), Bournemouth Questionnaire, satisfaction with the received care (7-point scale), EuroQol Questionnaire and last of all, standardized face-to-face interviews. Self-report questionnaires were completed at each time point, independent from study providers and investigators.

Measurement protocol

Instrument

Kinematics of the lumbar spine were sampled using a six-degrees-of-freedom instrumented spatial linkage system with a sampling rate of 100 Hz (CA 6000 Spine Motion Analyzer; OSI, Union City, CA, USA). The system is an electropotentiometric goniometer system in which six potentiometers are used (Figure 3). The potentiometers interpret positional change and displacement through change in voltage resistance. The data that are received by the potentiometers are interpreted as angular and translational motion between two rigid bodies in a given timeframe. This enables 3D motion to be computed [155;188;189]. Figures 4 and 5 illustrate the use of the CA 6000. Each time the CA 6000 software was launched (at the beginning of each test day), the linkage unit was calibrated against a calibration bar and a zero-setting was performed for each participant in the neutral position before the first test.



Figure 3: Illustrating the CA6000 Spine Motion Analyzer attached to a person in neutral position and the accompanying computer.

The CA 6000 Spine Motion Analyzer has previously been verified for precision and accuracy and most studies have reported the device to have good accuracy and precision [172;189-193]. Christensen concluded that the instrument has a very high movement precision ($\pm 0.1^\circ$) for all six motion directions. However, the accuracy relative to manual protractors ranged from 2.0% to 11.5% and was considered less than acceptable [172]. For the sagittal plane, angular precision was found to be within $\pm 0.1^\circ$ with an accuracy ranging from 2.0- 6.0% [172]. However, in other studies, the lack of accuracy for the CA 6000 has not been confirmed. For example, Feipel et

al. reported that precision was $\sim 0.3^\circ$ and accuracy $\sim 1^\circ$ [194] and McGregor et al. tested the accuracy between the CA 6000 and an engineering mill (which may be considered more appropriate than comparing it to a protractor) and reported excellent accuracy. Therefore the current literature is not unequivocal for accuracy but the majority of the studies indicate that it is accurate; however, all studies indicate that it is precise.

Attachment and recording procedure

Lumbar dynamic motion was assessed using a stringent standard protocol developed by the factory providing the CA 6000. This included a detailed description of how to handle the equipment and the accompanying computer program.

Each participant wore a loose T-shirt and trousers. The instrument was attached as described in the factory protocol with the patient standing in a neutral position with relaxed arms hanging down by their sides. The fixed extremity of the linkage was mounted on the sacral crest (S2) using the manufacturer-supplied belt. The mobile end was mounted at the level of T7 using the original chest harness and the top edge of the horizontal metal pieces was aligned evenly with the inferior angles of the scapulae (which are level with the T7 spinous process). The pelvic harness was applied so that the binding posts were level with the posterior superior iliac spines. The neutral position was defined as the patient standing with eyes open, facing forward, with the feet positioned a shoulder width apart and arms hanging freely at the side, with the low back in a comfortable position. Each patient then performed several trial runs as a 'warm up'.



Figure 4: The CA6000 Motion Analyzer with a person in neutral, extension and flexion positions

Four types of motion were tested following the same sequence i.e. backward and forward bending (extension/flexion) (Figure 4), left and right turning (rotation), left and right bending (lateral flexion) and rolling the back in both directions (left/right circumduction) (Figure 5). For all test directions, stringent test instructions were verbally explained to the patients.

For backward and forward bending each patient received the following verbal explanation . *“Ok, I’ll have you find a neutral position for your low back. Place your arms across your chest and bend backwards from the waist as far as you can go. As you return to neutral, move your palms to your thighs and while sliding your palms down your legs, bend forward from the waist as far as you can go, and then return to neutral (arms across chest). It should be done at your own pace and without pausing”*.

For circumduction motion (full turning of the back), each patient received the following verbal explanation. *“Find your neutral position and look forward with your hands on your hips. First bend backwards, then roll to your left, forwards, to your right, to the back and return to neutral. It is important to go as far as you can go in all directions. This entire movement should be done at your own pace without pausing. So, it should look like this”*. After these trials the patients were asked to circumduct their back in the opposite direction and each patient received the following verbal explanation. *“Ok, this is the last one. It is the same as the one you just did, but you’ll go in the opposite direction. So go backwards, right, forwards, left, back and then back to neutral”*.

The general concepts were finding the neutral position and then moving in a given direction, where the participant was specifically instructed to go ‘as far as you can go’ and that ‘it should be done at your own pace and without pausing’. This sample of chronic LBP patients was not highly incapacitated and all participants could complete the test procedure i.e. pain was not a major limiting factor for this study sample. All patients received exactly the same instructions word for word. No specific instructions were given encouraging them to continue the testing despite feeling pain or to go through/pass the pain. In these instructions pain was not considered i.e. if a patient asked what if I feel pain; the examiner would repeat the instruction go ‘as far as you can go’. The patient group recruited had chronic mechanical back pain i.e. pain could be reproduced by back movements and the pain was generally persistent. For instance, when the patients were asked the following question; *“Counting back from today, how many weeks/months/years in a row have you experienced at least some back pain?”* more than 80% answered more than 1 year. Therefore it was accepted that patients experienced pain during the procedure. However all patients could complete the objective movement test procedure although they had some level of pain during some of the movements.

Data from successive trials were obtained in each test session until reaching a ROM difference of four degrees or less between subsequent trials. A maximum of six trials was allowed. No rest period was given between trials. The time of day was recorded but not taken into account in the planning of the two visits. The testing time duration for the complete protocol was approximately 10 minutes and included both sagittal and coronal plane motions as well as rotation and circumduction.



Figure 5: A sequence of pictures (read from top left to bottom right) illustrating the CA6000 Spinal Motion Analyzer attached to a person performing circumduction motion - neutral, extension, left lateral flexion, flexion, right lateral flexion, extension and back to neutral position.

McGregor et al. evaluated the possibility of errors induced by movement of the CA 6000 harnesses and found it to be minimal in all planes of motion ($< 0.05\text{mm}$) [143]; however the study sample for this evaluation was not specified, nor was the way in which the assessment was carried out. For the current study no measurements were done in order to determine if the fixator straps moved with or after regional lumbar motion recording.

Management of kinematic data

Spinal motion 1.0

A custom-made MatLab program was used to reduce the 3D motion data into single numbered motion parameters. A trial selection model was developed in order to choose and extract the estimated two best data trials from each set. The selection process was based on a computerized comparison of all movement trials in each set. The specific computerized selection criteria were: uniformity of curves, ROM, and length of curves. In addition, a manual visual inspection of all sagittal motion plots was done in conjunction with the computerized selection process. ROM parameters were developed relative to each motion test i.e. sagittal plane extension/flexion, frontal plane lateral flexion, axial rotation and circumduction motion.

The background for the developed regional lumbar motion parameters

Routine clinical examination of the spine involves an overall assessment of the patient's ability to move the spine including end ROM and different patterns such as smoothness, coupled motion and uniformity of e.g. side bending and may be more quantifiable than the assessment of the patient's perception of e.g. pain. Clinicians are typically using a rough visual inspection or simple measurements such as finger-to-floor distance to evaluate the previously mentioned very complex motion characteristics. Sometimes a patient may move in a slow and jerky manner but still reach a high end ROM (e.g. touch the floor when forward bending); whereas other patients may move smoothly but be limited in ROM. Although reliability and validity and indeed the usefulness of these approaches can be questioned, it is the clinical experience of the author of this thesis that changes in these motion patterns are often observed following treatments of LBP patients. Similar observations have been described by McGregor et al. who report that acute LBP patients exhibited a 'stepped' flexion extension motion [143]. The clinical experience of the author suggests that changes in these motion patterns do occur following manual treatments and are often linked with the patient's perception e.g. the patients report that they feel they move more freely. But how can these patterns be quantified? This study aimed to examine specific motion parameters that were considered likely to capture these proposed characteristics. Based on the literature searches, common motion parameters derived from real-time computerized instruments like CA 6000 are ROM and velocity, making these obvious choices in order to be able compare with those reported in the literature. Velocity has been proposed to be a more sensitive parameter than ROM [143]; however it does not capture the previously mentioned motion patterns. Therefore, we also developed new motion parameters that might better reflect (i) the patient's motion patterns in terms of smoothness of the motion (Jerk Index), and (ii) functional ability via a phase-plot area, which is a parameter named using all available planer data combining ROM and velocity during extension and flexion motion. In addition, circumduction areas were calculated using all 2D or 3D data points

measured during the circumduction motion and may therefore represent a more relevant measure than single-plane ROM when quantifying functional impairments.

Motion parameters

The following motion parameters were selected for further analysis (Manuscripts II, III, IV):

1. *ROM (degree)* was calculated as the total angular range of regional lumbar motion in the sagittal plane from maximum extension to maximum flexion (Figure 6).
2. *Maximum flexion velocity (degree/sec)* was calculated as the peak angular speed in the forward bending motion reached from full extension to full flexion (Figure 6).
3. *Mean flexion velocity (degree/sec)* was calculated as the average angular speed from maximum extension to maximum flexion (Figure 6).
4. *Maximum extension velocity (degrees/sec)* was calculated as the peak negative angular speed reached in the ROM from full flexion and back to the neutral position (Figure 6).
5. *Phase-plot area (degree²/sec)* was defined as the area comprised by the phase-plot of sagittal flexion-extension angular motion versus velocity. The phase-plot area was calculated based on cross-product calculations between vectors drawn from the neutral position (0,0) to each coordinate point in the phase-plot (Figure 7).
6. *Jerk Index* was calculated from maximum extension to maximum flexion as the mean spectral frequency of the first derivative of the angular acceleration signal multiplied by movement duration. This parameter indicates the number of changes in acceleration, i.e. the smoothness of the motion.

$$Jerk\ index = mpf \left(\frac{dA}{dt} \right) * t$$

where A = acceleration, mpf = mean power frequency, and t = duration of the movement.

7. *2D circumduction area (degree²)* was defined as the 2-dimensional (2D) surface area of the angular phase-plot formed by the frontal and sagittal motion. The area was calculated based on cross-product calculations between vectors drawn from the neutral position (0,0) to each coordinate measurement point in the circumduction motion. The average of left and right circumduction areas was used in the analysis (Figure 8).
8. *3D circumduction area (cm²)* was defined as the curved 3D surface formed by the translatory motion. The area was calculated based on cross-product calculations between vectors drawn from the neutral position (0,0,0) to each (x, y, z) coordinate measurement point in the circumduction motion. The average of left and right circumduction areas was used in the analysis.

For practical reasons, we did not include all regional lumbar motion parameters in each manuscript. For the reliability study (Manuscript II), we decided to evaluate all sagittal motion parameters only. For the following studies (Manuscripts III and IV), we included the circumduction motion parameters, and in order to have a manageable number of parameters, we dropped the velocity parameters of maximum extension velocity and mean flexion velocity in these analyses.

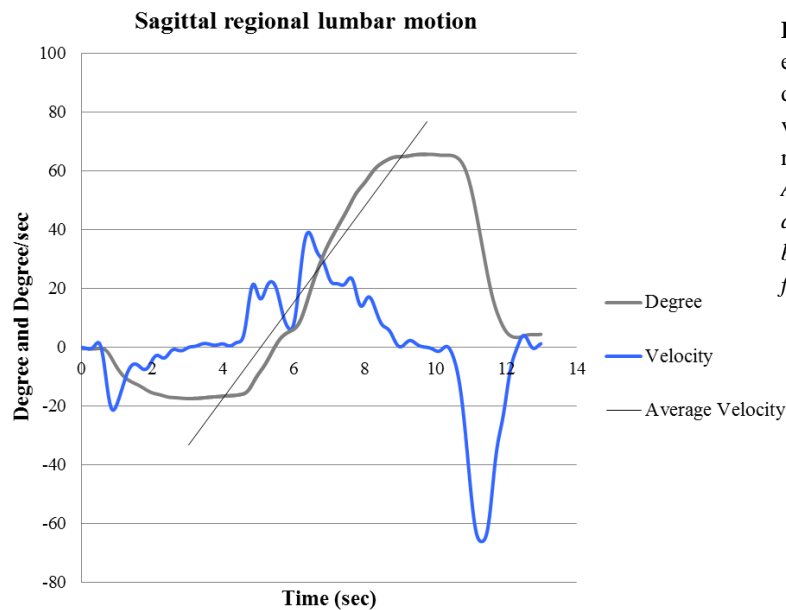


Figure 6: Regional lumbar extension and flexion motion displayed in degrees and velocity at a single representative session. *Note: Average velocity was measured as the slope of the straight line between maximum extension and flexion positions.*

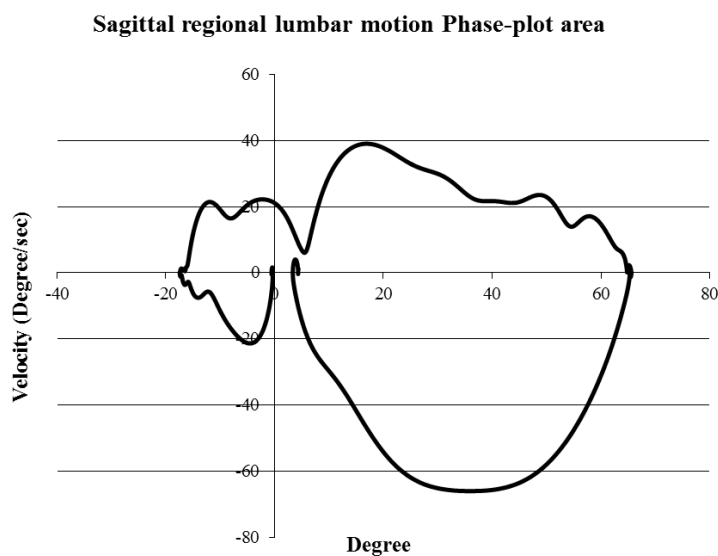


Figure 7: Regional lumbar extension and flexion motion obtained in a single representative subject, displayed as a position (degrees) - speed (velocity) phase-plot. The area within the two curves is defined as the spinal motion area. *Note: Movement was started at (0,0) from which the subject performed an extension motion to reach maximum extension ROM (negative position and velocity values) followed by a flexion motion to reach maximum flexion ROM (positive velocity values) and subsequently returned to neutral position (negative velocity values, positive position values).*

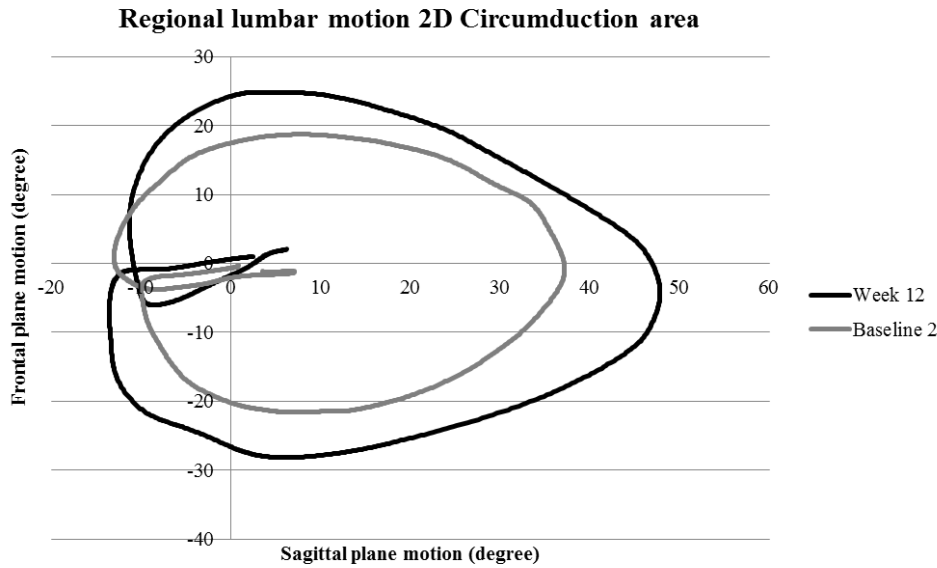


Figure 8: Regional lumbar circumduction motion in a typical patient before and after treatment. The area increased after treatment.

Statistical analysis

Shared statistical analysis for Manuscripts II, III and IV

Paired t-test or Wilcoxon signed rank test were used for comparison between paired data, and Wilcoxon Rank-Sum (Mann-Whitney U) test, Chi²-test and unpaired t-tests were used for comparison of unpaired data.

Statistical analyses were carried out in order to determine if those participants who completed the motion tests were different to those who did not. For the baseline analysis in Manuscript II (n=301 vs. 220) and the follow-up analyses in Manuscripts III and IV (n=301 vs. 199), independent t-tests or chi-square tests were calculated between groups on the following parameters: age, sex, BMI (kg/m²), duration of pain, baseline physical component score (SF-36), baseline mental component score (SF-36), baseline depression score (CESD), Quebec Task Force classification, LBP intensity and leg pain intensity (ordinal 11-box scale) [186], RMDQ [187] and intervention group.

The assumption of normally distributed and homoscedastic data was tested in a variety of ways including a Q-Q plot on Bland-Altman plots. In addition, the Shapiro-Wilk W test was performed for the residuals in order to check for normal distribution.

For the follow-up analyses (Manuscripts III and IV), all analyses were based on change scores between the second baseline and after 12 weeks of intervention.

Stata 10.1 or 11.1 (Statacorp, College Station, TX, USA) were used for all analyses.

Specific statistical analysis for the reliability study on chronic LBP patients (Manuscript II)

All lumbar motion parameters except for ROM were non-normally distributed. The Shapiro-Wilk W test and the various plots indicated a better fit with the statistical assumption after natural log transformation (Figure 9). Therefore, except for the ROM motion parameter, all analyses were based on natural log transformed data.

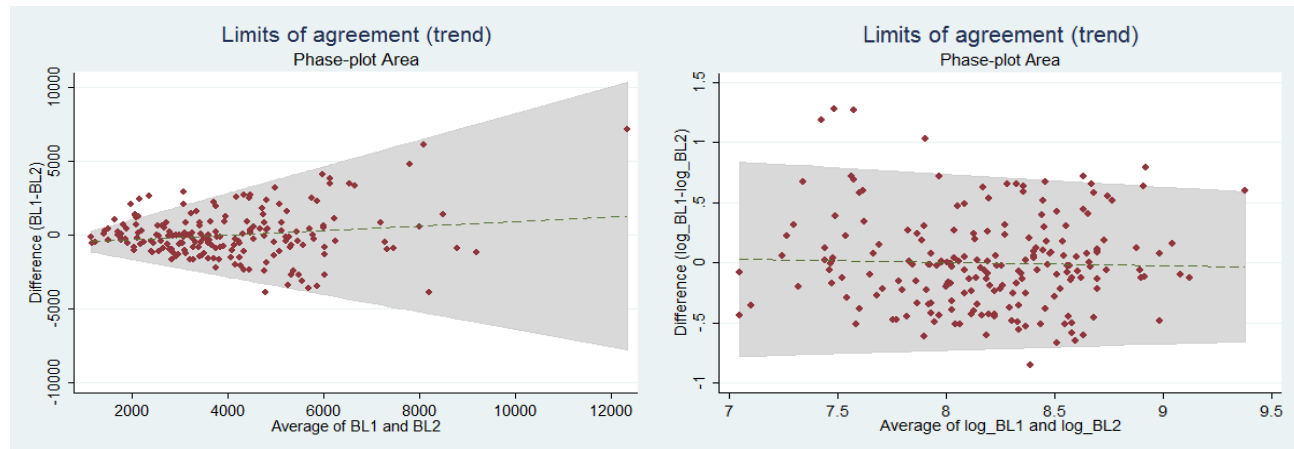


Figure 9: Illustration of a Bland-Altman plot adjusted for trend. A regression model is used to adjust the 95% limits of agreement indicating a more normally distributed pattern and that the heteroscedastic appearance was corrected when using natural log transformation. The parameter illustrated in these examples was the Phase-Plot motion at baseline 1 and 2.

A number of statistical tools were used to assess test-retest reliability and measurement error: Paired t-testing was used to detect systematic bias between test sessions. Based on the study design, $ICC_{(1,1)}$ were calculated to assess reliability [195]. To assess measurement error, LOA with 95% confidence intervals (CI) were calculated [171].

Stratification

Subgroup stratification based on pain distribution, (back pain only versus back and leg pain) was dichotomized by combining Quebec diagnostic groups 2, 3 and 4 versus Quebec diagnostic group 1. Stratification into subgroups concerning BMI was based on the cut-off points proposed by the WHO for the classification of overweight i.e. a BMI greater than or equal to 30 kg/m^2 [196]. The subgroup, including patients with an unstable pain level, was defined as patients with a change in the VAS score between test and retest of ± 2 or more [197].

Specific statistics for analysis of lumbar motion changes (Manuscript III)

Paired t-test or Wilcoxon signed rank test were used for comparison between paired data, and Wilcoxon Rank-Sum (Mann-Whitney U) test was used for comparison of unpaired data. For comparison of differences in pre-to-post changes between the three treatment groups, the Kruskal-Wallis test was used.

Stratification

Analysis was done relative to the different treatments (SMT, SET, or HEA over a 12-week period).

Specific statistics for analysis of the association between regional lumbar motion and patient-rated outcomes (Manuscript IV)

No significant difference between the three treatment groups SET, SMT and HEA in terms of pain and other patient-rated outcomes, in short- and long-term were found in the primary analysis [1]. Based on these results we found it acceptable to collapse these treatment groups in order to analyses associations in changes in pain and back-related function (RMDQ) versus regional lumbar motion in the total cohort.

For identifying clinically meaningful improvements in the measurement of back pain and back-related function, we used a 30% threshold as recommended by Ostelo et al. [198]. An increase in motion score was considered to be related to an improvement in patient-rated outcome with the exception of the Jerk Index, which was expected to be the opposite i.e. clinical improvement would result in a smoother motion and, thus, a lower score.

All regional lumbar motion parameters except for ROM were non-normally distributed. Therefore we presented mean values for ROM but median values of all other regional lumbar motion parameters with 95% confidence intervals and calculated Spearman correlation coefficients (Table 2,3,4,5 and 6, Manuscript IV). Patients who became worse (deteriorated) is also presented in tables but the groups were very small (ranging from 1 to 8 participants).

Stratification

In order to examine how this potential relationship relates to other factors we also did an exploratory analysis in order to investigate relationships of the data relative to subgroups i.e. based on pain distribution, (back pain only versus back and leg pain), and different treatments (SMT, SET and HEA).

Summary of results

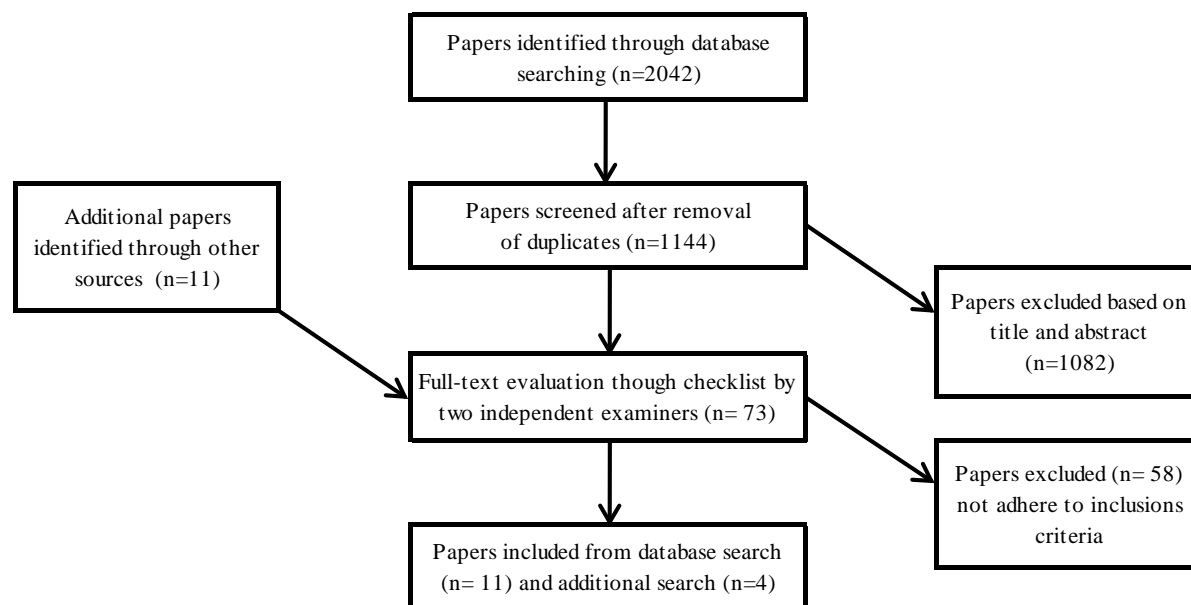
The results of the review (Manuscript I), and the cohort studies (Manuscripts II (reliability study), III and IV (intervention studies)) will be reported separately.

Results of the literature review

Included articles

In total, the titles and abstracts of 2,042 papers were retrieved and examined according to the inclusion criteria. After careful screening, 15 of those that were retrieved in hard copy were retained for inclusion (Figure 10). For the included studies, reproducibility or reliability of spinal lumbar motion was reported as a the sole investigative target in seven of them [156-158;189;199-201], while the remaining eight studies investigated reproducibility or reliability as part of a study addressing another primary aim[58;155;159;188;192;193;202;203]. There were only minor disagreements between the two reviewers on the selection of the studies, extraction of data or quality assessment score. All disagreements were easily resolved by discussion.

Figure 10: Flow diagram illustrating study selection



Descriptive items

Overall study information

Detailed descriptive information on all included studies is provided in Tables 2, 3 and 4 (Manuscript I). In total, five different 3D motion instruments were examined.

Study population

In general, the level of reporting regarding study populations was incomplete (Tables 2, 3 and 4, Manuscript I). Altogether, a total of 132 men and 129 women, plus 34 subjects of unknown gender, participated in the 15 studies, with the average number of participants being 20 (range: 6 to 31). A considerable difference in the detail of reporting of age and anthropometric data between the studies was observed. The age ranged from 20 to 72 years with a mean of 27.5 ± 8.3 SD and the BMI ranged from 21 to 27 with a mean of 23.9 ± 1.8 SD, based on the available information (Tables 2, 3 and 4, Manuscript I). In seven studies, no anthropometric information was reported [155;156;158;188;192;200;202]. Specific inclusion/exclusion criteria were described in only five of the 15 articles [159;188;193;199;203]. Twelve papers included pain-free subjects and three included patients with LBP [193;199;200].

Testing circumstances

The level of reporting regarding the testing protocol varied considerably between studies and was mostly incomplete (Tables 2, 3 and 4, Manuscript I). The educational background of the examiners was described in five of the 15 studies [159;189;192;201;202] (Tables 2, 3 and 4, Manuscript I), and the experience of the examiners in working with elements of the protocol was reported in five articles e.g. 'familiar with skin surface marking techniques for spine', 'trained in use of equipment', 'four hours training with the equipment' [159;189;192;201;203]. Information on blinding of examiners was provided in three of the 15 articles [155;159;188]. The time interval between test and retest varied from approximately one month [58] to one or two weeks [159;188] to days; [155;156;192;201-203] to the same day [157;159;189;193;199] and was not reported in two studies [158;200].

Equipment

In general, the description of the equipment used was complete when provided; however a description of instrument accuracy and/or precision was not provided in nine papers [155;156;158;159;188;193;201-203]. The instrument outcome reported was ROM in 13 studies, [58;155;156;159;188;189;192;193;199-203] ROM and higher order kinematics in four, [58;156;192;200] and ROM and motion patterns in one [157]. In one study, no outcome measure was reported [158].

Data analysis and presentation

Data presentation was sufficient to assess analysis adequacy in four of the 15 studies [156;189;192;200].

A range of data analyses and statistical methods were applied in the papers addressing reliability (Tables 2, 3 and 4, Manuscript I). ICCs were reported with formulae specified in four studies [159;189;193;203]. The ICCs were reported without the formulae specified in four studies [156;192;199;201]. Cronbachs alpha was used in one study,[58] correlation coefficients were used in three studies [155;157;188] and in another study, no statistical methods were reported [158].

Different methods were also used in the reporting of agreement parameters. The Bland and Altman mean difference technique was applied as mean difference \pm SD in one study[192] and was applied with different criteria (mean difference \pm 2SD) in another [200]. The standard error of measurement was reported in four studies [159;189;193;202] and the coefficient of variation in three studies [155;188;202].

Quality assessment

Figure 11 contains the summarised quality assessments of all articles relative to each domain. For the domains ‘study population’ and ‘testing circumstances’, the articles that scored ‘yes’ for complete or near-complete reporting, were only 33% and 20% respectively. The domain ‘equipment’ had the highest completeness of reporting with 73% of the articles meeting this criterion. Only 13% of the articles were assigned ‘yes’ for the reporting of ‘data presentation’. Table 2 contains the studies and their individual quality assessments grouped by instrument.

Figure 11: Overview of results of the quality assessment in each domain

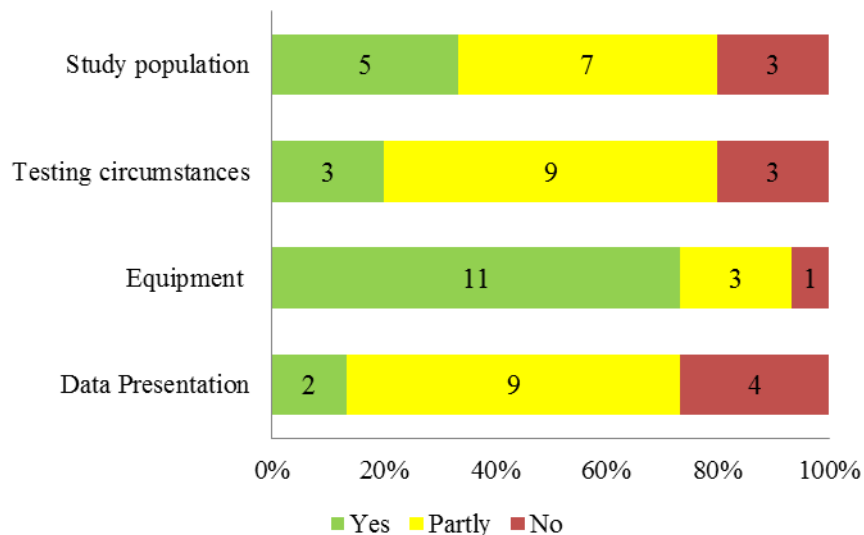


Table 2: Overview of all included studies and its individual quality ratings yes, partly and no

Quality assessment	Study population	Testing circumstances	Equipment	Data presentation
<u>CA 6000</u>				
Dopf (1994)	Partly	Partly	Partly	Partly
Dvorak (1995)	No	Partly	Yes	Partly
McGregor (1995)	Partly	Partly	Yes	Partly
McGregor (2000)	Partly	No	Yes	Yes
Petersen (1994)	Partly	Partly	Yes	Yes
Petty (1995)	No	Partly	Partly	No
Schuit (1997)	Yes	Partly	Yes	Partly
<u>Modified CA 6000</u>				
Troke (1996)	Yes	Yes	Yes	Partly
Troke (2007)	Yes	Partly	Yes	Partly
<u>3Space Fastrak</u>				
Barret (1999)	Partly	Partly	Partly	No
<u>Lumbar motion monitor</u>				
Gill (1996)	Partly	Yes	Yes	Partly
Marras (1995)	Yes	No	Yes	Partly
<u>Zebris CMS20</u>				
Moutzouri (2008)	Yes	Yes	Yes	Partly
<u>Mac Reflex</u>				
Vanneuville (1994)	No	No	No	No
<u>IS 300 Pro. gyroscope</u>				
Lee (2003)	Partly	Partly	Yes	No

Reliability

ICC values for ROM parameters were found to be 0.75 or greater in all motion directions in six studies using intra-tester designs [156;159;189;193;199;201] and in three studies using inter-tester designs [156;189;201]. In two studies, lower coefficients were found in some of the motion directions [192;203].

Two studies examined higher order kinematics [156;192]. McGregor et al. reported ICC velocity parameters ranging from 0.67 to 0.86 for intra-tester repeatability and from 0.74 to 0.98 for inter-tester agreement [192]. Gill and Callaghan found ICC values ranging from 0.61 to 0.87 for intra-tester repeatability and 0.93 or more for inter-tester agreement [156]. In addition, Gill and Callaghan reported acceleration parameters ranging from 0.46 to 0.72 for intra-tester agreement and from 0.95 to 0.97 for inter-tester repeatability [156].

Agreement

Standard error of measurement values based on ROM data were presented in three studies [159;189;193] ranging from 2.3° to 6.5° and were reported as root mean square error (7.43-8.6°) in one study [202]. The Bland and Altman mean difference LOA technique was reported in two studies based on ROM and velocity data [192;200]. Intra and inter-tester (mean difference \pm SD) values based on ROM values ranged from -1.8° to 2.4° (\pm 2.4 to 5.9°) and velocity values ranged from 0.1 to 3.8°/sec (\pm 5.3 to 15.4°/sec) [192]. In the other study, intra and inter-tester (mean difference \pm 2 SDs) values based on ROM values ranged from -1.8° to 2.2° (\pm 5 to 14.2°) and velocity values ranged from -49.4 to 8.3°/sec (\pm 7.8 to 131.0°/sec) [200].

Results of the cohort studies

Study sample

A total of 630 individuals were evaluated for the study, of which 329 were excluded because they did not meet the exclusion criteria specified in the primary paper [1]. Therefore, 301 patients were recruited, but due to technical problems with the equipment, a total of 220 complete patient recordings were obtained for Manuscript II and 199 for Manuscripts III and IV. The individuals not available for analyses were younger (significantly for manuscript III and IV) but there were no differences in other baseline characteristics such as BMI, gender, duration of pain, or depression score, back/leg pain intensity and RMDQ score (Table 3). For Manuscript II, 59 individuals were not included because of missing data or technical problems with the recordings on one of the test days, plus 22 dropouts, and for the follow-up manuscripts (III and IV), 80 patients were not included for the same reason, plus 22 dropouts. Of the participants available for Manuscripts III and IV, 62 received SET, 77 received SMT and 60 received HEA. Descriptive data of the cohort are also presented in Manuscripts II, III and IV. Overall, adherence to study interventions was high i.e. the number of patients who did not receive or discontinued intervention for each treatment group were: 4 for HEA (refused to participate n=3, time commitment n=1), 4 for SMT (refused to participate n=3, competing co-morbidity n=1) and 14 for the SET (Unknown reason n=2, increase in pain n=3, refused to participate n=3, moved n=1, time commitment n=1, personal conflict n=3). More detail about the adherence is reported in the primary paper [1]. Table 3 summarizes the demographic and clinical characteristics of the study participants and the participants not available for analysis.

Table 3: Patient characteristics

Characteristic	Baseline		Follow-up		Drop out	
	MD	NMD	MD	NMD		
	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)
Number of subjects	220	59	199	80	22	
Age (years)	45.5 (11.2)	44.6 (10.4)	45.9 (11.3)	44.0 (10.0)	41.7 (10.5)	
BMI	28.1 (5.6)	27.7 (5.4)	28.1 (5.6)	27.7 (5.5)	27.2 (3.8)	
Males (%)	40	42	40	40	36	
Activity level	2.4 (1.2)	1.9 (1.5)	2.4 (1.2)	2.1 (1.5)	2.0 (1.5)	
Past episodes of LBP	3.3 (0.8)	3.2 (0.7)	3.2 (0.8)	3.2 (0.7)	3.2 (0.8)	
Pain more than 1 year (%)	81	85	82	83	91	
<u>Quebec Task Force Classification (%)</u>						
1	68	64	67	67	55	
2	18	19	18	18	32	
3	12	10	13	10	5	
4	2	7	2	5	9	
<u>Back pain level (0-10)</u>						
Baseline 1	5.2 (1.6)	5.4 (1.6)	5.1 (1.5)	5.5 (1.7)	6.0 (1.7)	
Baseline 2	5.2 (1.7)	5.1 (1.6)	5.2 (1.7)	5.2 (1.7)	6.0 (1.8)	
Week 12	2.8 (2.0)	2.9 (2.0)	2.8 (1.9)	3.1 (2.3)	4.0 (3.5)	
<u>Leg pain level (0-10)</u>						
Baseline 1	2.7 (7.0)	1.6 (2.1)	2.8 (7.3)	1.6 (2.2)	1.9 (2.6)	
Baseline 2	2.0 (2.5)	1.9 (2.4)	2.1 (2.5)	1.9 (2.4)	2.1 (3.1)	
Week 12	0.9 (1.7)	1.0 (1.8)	0.9 (1.6)	1.1 (1.8)	0.6 (1.3)	
RMDQ baseline 1	8.5 (4.3)	9.7 (5.2)	8.4 (4.4)	9.5 (4.9)	9.4 (4.7)	
RMDQ baseline 2	8.5 (4.9)	8.7 (4.9)	8.4 (4.9)	8.6 (4.8)	8.4 (5.4)	
RMDQ week 12	18.7 (20.3)	19.3 (21.8)	18.5 (20.4)	19.5 (20.9)	25.6 (21.5)	
CESD baseline 1 §	12.6 (10.4)	15.7 (10.4)	12.4 (10.4)	15.1 (10.1)	17.1 (12.9)	
CESD baseline 1	11.2 (9.8)	13.5 (10.7)	11.1 (10.0)	13.1 (10.2)	13.8 (11.7)	
CESD week 12	8.5 (10.2)	9.0 (7.5)	8.3 (10.2)	9.6 (8.3)	9.3 (9.0)	

(MD) = motion data available, (NMD) = no motion data available, Activity level: Engaged in exercise or sports activities in the past month? (0 = I do not engage in exercise or sports, 1 = Less than once a week, 2 = Once a week, 3 = 2 or 3 times per week, 4 = 4 times or more per week), Past episodes of LBP: An episode is a week with at least some LBP (0 =None, 1 = 1-2 episodes, 2 = 3-5 episodes 3=More than 5 episodes, 4= 1 single episode of continuous LBP), Pain more than 1 year: information gained through the following question “Counting back from today, how many weeks/months/years in a row have you experienced at least some back pain?”, Quebec Task Force classification system: Group 1 = LBP patients without radiation, Group 2 = LBP patients with pain radiation to proximal lower extremity, Group 3: LBP patients with pain radiation to distal parts of lower extremity, Group 4 = LBP patients with pain radiation to lower extremity accompanied by neurological signs, Pain level (0-10) measured on a ordinal 11-box pain scale with 10 being the worst possible pain, (CESD) = Center for Epidemiologic Studies Depression Scale, § Depression is defined as greater than 16 points on the Center for Epidemiologic Studies Depression Scale (CESD).

Reliability and measurement error

Only sagittal motion parameters have been assessed in Manuscript II; however in this section, we also report on the two circumduction motion parameters. All regional lumbar motion parameters except total sagittal ROM showed a heteroscedastic appearance when displayed in Bland-Altman plots, implying that the magnitude of measurement error was associated with the magnitude of the parameter [171]. When these measurements were log transformed, the mean and SD of the test-retest differences remained more similar throughout the range of parameter values, indicating that this procedure would provide a better fit to the statistical model. The Shapiro-Wilk W test calculated on the residuals indicated a more normal distribution of the data when log transformed, except for the ROM parameter. Therefore $ICC_{(1,1)}$ and Bland-Altman LOA were calculated with logarithmically transformed data, except for total ROM.

We hypothesized that reliability for all motion parameters would have an $ICC_{(1,1)}$ in the range of (0.6-0.8) in participants with a stable pain level. The $ICC_{(1,1)}$ ranged from 0.55 to 0.79 (Table 5). Therefore, with the exception of the Jerk Index ($ICC_{(1,1)} = 0.55$) the regional lumbar motion parameters reliability coefficient were within the hypothesised range, indicating moderate reliability (0.5-0.75) [168].

Overall, a statistical difference was observed between baseline measurement session 1 and session 2 for the three motion parameters of mean flexion velocity (16% higher, $P=0.001$), maximal flexion velocity (6% higher, $P=0.011$) and Jerk Index (22% lower, $P=0.001$). All other parameters showed no systematic difference between sessions 1 and 2. $ICC_{(1,1)}$ values for the motion parameters calculated using the total LBP group ranged between 0.51 to 0.81 and wide LOA were observed for all parameters (Table 4).

Reliability measures ($ICC_{(1,1)}$) in patient subgroups ranged between 0.34 and 0.85 (Table 5). For the majority of the sagittal motion parameters, higher $ICC_{(1,1)}$ coefficients and smaller LOA were found in subgroups with patients examined by the same assessor, patients with stable pain level, patients with BMI below 30 kg/m², male patients, and patients in the Quebec Task Force classification group 1. However, for the circumduction parameters, the same pattern only existed when stratifying into examiner and gender (Table 5).

Table 4: Reliability and measurement error of lumbar regional lumbar motion parameters

Parameter (n=220)	ICC _(1,1) (CI 95%)		95% LOA
ROM (degree)	0.69	(0.62-0.76)	(-23-27)
Flexion mean velocity (degree /sec)	0.61	(0.53-0.70)	(0.4-1.83)
Extension max velocity (degree /sec)	0.7	(0.63-0.76)	(0.55-1.71)
Flexion max velocity (degree /sec)	0.64	(0.56-0.72)	(0.5-1.79)
Phase-plot Area (degree ² /sec)	0.69	(0.62-0.76)	(0.47-2.11)
Jerk Index	0.51	(0.42-0.61)	(0.57-2.59)
2D (degree ²)	0.81	(0.76-0.86)	(0.59-1.72)
3D (cm ²)	0.68	(0.61-0.75)	(0.60-2.48)

ICC_(1,1) = Intraclass correlation coefficient, LOA = Limits of agreement, ROM = Range of motion,, Jerk Index = number of changes in acceleration from full extension to full flexion, 2D = two dimensional circumduction area, 3D= three dimensional circumduction area, *All parameters except ROM were natural log transformed to fit the statistical model and are therefore presented in LOA (ratio).

Table 5: Regional lumbar motion reliability and measurement error for LBP patients divided into subgroups

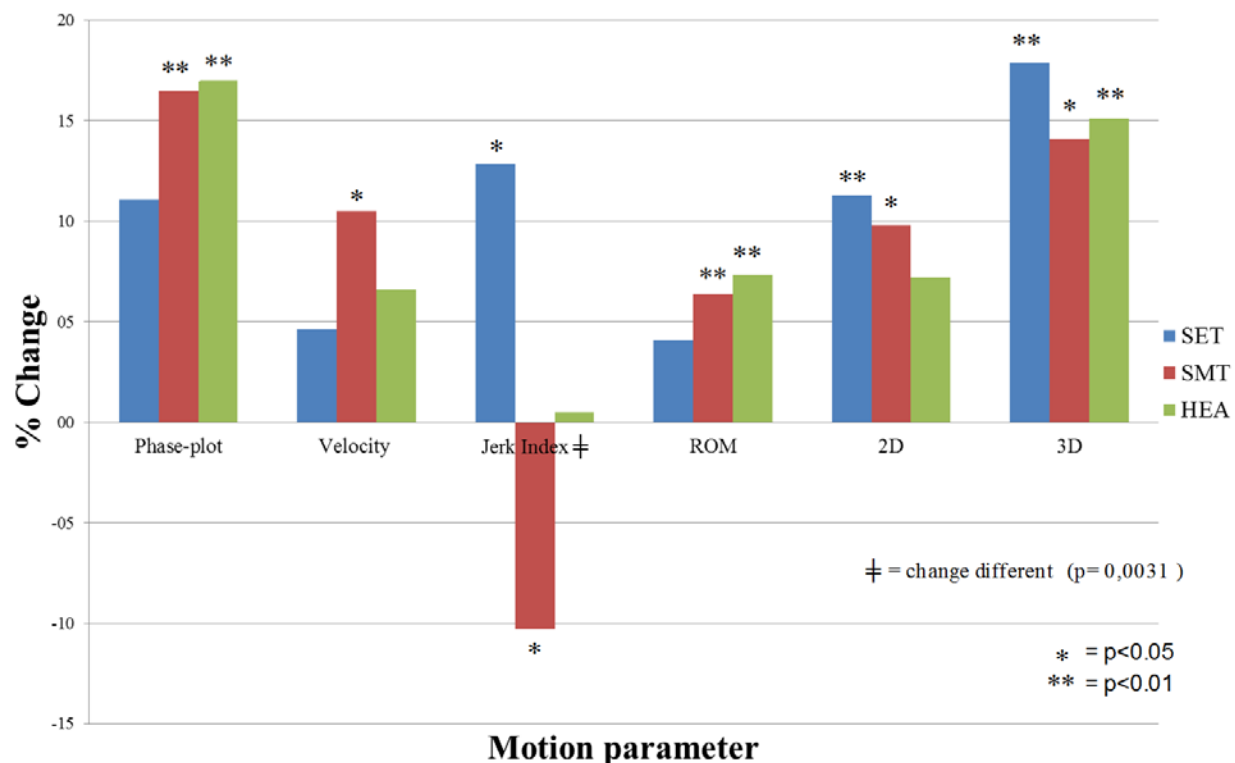
Motion Parameter	Statistical parameter	Subgroups									
		Same ex.	Different ex.	BMI <30	BMI >30	Pain (s)	Pain (u)	Male	Female	Group 1	Group 2,3,4
	Number of subjects	89	131	147	73	162	58	87	133	149	71
ROM	ICC _(1,1)	0.77	0.64	0.66	0.73	0.69	0.7	0.66	0.65	0.68	0.69
	LOA _{UL} (degree)	22	30	26	27	26	29	24	28	26	28
	LOA _{LL} (degree)	-20	-25	-24	-21	-22	-28	-25	-22	-25	-21
Mean velocity	ICC _(1,1)	0.54	0.65	0.63	0.58	0.64	0.53	0.7	0.53	0.65	0.55
	LOA _{UL} (ratio)	1.79	1.86	1.8	1.89	1.71	2.18	1.66	1.95	1.71	2.1
	LOA _{LL} (ratio)	0.39	0.4	0.42	0.36	0.41	0.37	0.42	0.39	0.41	0.38
Max flexion velocity	ICC _(1,1)	0.67	0.61	0.67	0.58	0.69	0.48	0.72	0.57	0.66	0.59
	LOA _{UL} (ratio)	1.6	1.91	1.68	2.03	1.65	2.19	1.72	1.85	1.69	2.02
	LOA _{LL} (ratio)	0.49	0.51	0.51	0.48	0.51	0.47	0.52	0.48	0.51	0.48
Max extension velocity	ICC _(1,1)	0.74	0.67	0.75	0.6	0.71	0.63	0.71	0.68	0.69	0.7
	LOA _{UL} (ratio)	1.63	1.77	1.6	1.93	1.67	1.83	1.69	1.73	1.68	1.79
	LOA _{LL} (ratio)	0.57	0.53	0.58	0.5	0.55	0.55	0.53	0.57	0.56	0.53
Phase-plot Area	ICC _(1,1)	0.76	0.64	0.7	0.68	0.74	0.56	0.7	0.66	0.7	0.67
	LOA _{UL} (ratio)	1.87	2.27	1.95	2.41	1.92	2.65	1.87	2.26	1.92	2.5
	LOA _{LL} (ratio)	0.5	0.46	0.48	0.47	0.5	0.42	0.48	0.47	0.49	0.45
Jerk Index	ICC _(1,1)	0.50	0.52	0.56	0.41	0.55	0.42	0.70	0.34	0.55	0.41
	LOA _{UL} (ratio)	2.60	2.59	2.45	2.86	2.45	2.95	2.31	2.78	2.53	2.72
	LOA _{LL} (ratio)	0.58	0.56	0.58	0.56	0.61	0.48	0.61	0.55	0.57	0.57
2D	ICC _(1,1)	0.84	0.79	0.78	0.83	0.79	0.85	0.82	0.79	0.81	0.81
	LOA _{UL} (ratio)	1.58	1.82	1.73	1.71	1.69	1.81	1.56	1.83	1.66	1.82
	LOA _{LL} (ratio)	0.63	0.57	0.57	0.64	0.61	0.54	0.64	0.56	0.57	0.64
3D	ICC _(1,1)	0.80	0.61	0.64	0.73	0.66	0.73	0.67	0.63	0.65	0.72
	LOA _{UL} (ratio)	1.94	2.86	2.49	2.46	2.52	2.38	2.00	2.81	2.42	2.57
	LOA _{LL} (ratio)	0.48	0.38	0.40	0.45	0.42	0.40	0.47	0.39	0.39	0.48

Ex. = examiner(s), BMI = body mass index (kg/m²), Pain (s) = pain score max change ± 1 , Pain (u) = pain score change ± 2 or more, Group = Quebec Task Force classifications 1 vs. 2, 3 and 4, ROM= Range of motion (degree), Velocity = (degree/sec), Phase-plot = Phase-plot Area (degree²/sec), Jerk Index = number of changes in acceleration from full extension to full flexion, 2D = 2-dimensional circumduction area (degree²), 3D = 3-dimensional circumduction area (cm²), ICC_(1,1) = Intraclass correlation coefficient, LOA_{UL} = Limits of agreement (upper limit), LOA_{LL} = Limits of agreement (lower limit), NB: Max and min values in each row are bolded.

Regional lumbar motion changes

For the regional lumbar motion evaluation, 199 persons had complete motion data at baseline and Week 12. Of these, 62 received SET, 77 received SMT and 60 received HEA. We hypothesized that the groups receiving either SET or SMT care would change significantly in all motion parameters over a 12-week period, whereas the minimal intervention group (HEA) would not (no change in motion parameters). Specifically the Jerk Index was hypothesized to decrease and all other motion parameters to increase. The SMT group increased on all parameters except for the Jerk Index, which decreased significantly (Figure 12 and Table 6). The two exercise groups increased significantly on 3 out of 6 motion parameters. The pre-to-post change in Jerk Index differed between treatments ($p = 0.0031$), with the SMT group changing to a smoother motion. Therefore we could not confirm our hypothesis.

Figure 12: Lumbar motion percentage changes between baseline and 12-week follow up by motion parameters and treatment groups from data presented in table 6



SET = Supervised exercise therapy, SMT = Spinal manipulative therapy, HEA = Home exercise and advice, Phase-plot = Phase-plot Area ($\text{degree}^2/\text{sec}$), Velocity = Maximum flexion velocity (degree/sec), Jerk Index = number of changes in acceleration from full extension to full flexion, ROM = Range of motion (degree), 2D = 2-dimensional circumduction area (degree^2), 3D = 3-dimensional circumduction area (cm^2). ROM p values using paired t-test. All others using Wilcoxon signed rank test. ‡ change different for SMT vs. SET and HEA (Kruskal-Wallis test). NB: This figure is based on the data and statistics presented in Table 6. It is made for descriptive purposes and interpretations should be done carefully.

Table 6: Lumbar motion characteristics at baseline and 12-week follow up by treatment group

Supervised exercise therapy (SET)					
(n=62)	Baseline		Follow Up 12 Weeks		
Spinal Measure	Mean	(95% CI)	Mean	(95% CI)	p value
<u>Flexion</u>					
Phase-plot	4038	(3611 - 4465)	4485	(3956 - 5013)	0.102
Velocity	36.8	(33.8 - 39.7)	38.5	(34.6 - 42.5)	0.477
Jerk Index	10.1	(9.2 - 11.1)	11.4	(10.1 - 12.6)	0.021
ROM	71	(67.2 - 74.8)	73.9	(69.9 - 78)	0.092
<u>Circumduction</u>					
2D	2722	(2448 - 2996)	3029	(2704 - 3353)	0.01
3D	144	(125 - 163)	170	(145 - 195)	0.004
Spinal manipulative therapy (SMT)					
(n=77)	Baseline		Follow Up 12 Weeks		
Spinal Measure	Mean	(95% CI)	Mean	(95% CI)	p value
<u>Flexion</u>					
Phase-plot	4018	(3601 - 4434)	4680	(4175 - 5185)	0.003
Velocity	36.2	(32.8 - 39.6)	40	(36.3 - 43.8)	0.046
Jerk Index \neq	10.7	(9.5 - 11.9)	9.6	(8.7 - 10.6)	0.014
ROM	69	(65.5 - 72.5)	73.4	(69.4 - 77.4)	0.011
<u>Circumduction</u>					
2D	2714	(2442 - 2985)	2980	(2653 - 3307)	0.053
3D	135	(117 - 153)	154	(134 - 174)	0.028
Home exercise and advice (HEA)					
(n=60)	Baseline		Follow Up 12 Weeks		
Spinal Measure	Mean	(95% CI)	Mean	(95% CI)	p value
<u>Flexion</u>					
Phase-plot	3884	(3396 - 4372)	4544	(3966 - 5123)	0.005
Velocity	34.9	(31.6 - 38.2)	37.2	(33.9 - 40.6)	0.169
Jerk Index	9.9	(8.8 - 10.9)	9.9	(8.9 - 11)	0.988
ROM	68	(63 - 73)	73	(68 - 78.1)	0.005
<u>Circumduction</u>					
2D	2727	(2384 - 3070)	2924	(2540 - 3308)	0.069
3D	139	(116 - 162)	160	(134 - 187)	0.006

Phase-plot = Phase-plot Area (degree²/sec), Velocity = Maximum flexion velocity (degree/sec), Jerk Index = number of changes in acceleration from full extension to full flexion, ROM= Range of motion (degree), 2D = 2-dimensional circumduction area (degree²), 3D = 3-dimensional circumduction area (cm²). ROM p values using paired t-test. All others using Wilcoxon signed rank test. \neq change different for SMT vs. SET and HEA (Kruskal-Wallis test p= 0,0031).

Association between change scores in regional lumbar motion and patient-rated outcomes

Our hypothesis that patients who had a clinically relevant improvement in pain and back-related function (RMDQ) would have greater change scores in velocity, ROM, circumduction area, and have a smoother motion compared to patients who did not achieve a clinically relevant improvement, could not be confirmed. Some motion parameters however did change as hypothesized but the confidence intervals were wide i.e. none were significant (Table 7). In general, the participants who deteriorated, i.e. experienced at least 30% deterioration in back pain, decreased in all motion parameters and changed to a less smooth motion. However the same consistent patterns were not found between motion parameters and back-related function and in addition only 8 participants deteriorated.

Table 7: Regional lumbar motion changes vs. clinically relevant changes in all included patients

All	>30% Improvement in BPI (n=143)	no change in BPI (n=48)	>30% Deterioration in BPI (n=8)
Measure	Median (95% CI)	Median (95% CI)	Median (95% CI)
P-P	580 (330-850)	296 (-206-831)	-675 (-1349-2277)
Vel	2,3 (-0.2-4.3)	2,8 (-0.5-6.1)	-2,0 (-13.9-9.5)
Jerk	0,05 (-0.7-0.8)	-0,1 (-1.2-1.1)	1,2 (-3.2-9.5)
#ROM	4,8 (2.5-7.1)	2,8 (-0.9-6.6)	-0,7 (-15.2-13.8)
2D	210 (108-388)	197 (-110-365)	-402 (-1335-934)
3D	18 (8-29)	28 (7-48)	-20 (-112-74)

All	>30% Improvement in RMDQ (n=136)	no change in RMDQ (n=53)	>30% Deterioration in RMDQ (n=8)
Measure	Median (95% CI)	Median (95% CI)	Median (95% CI)
P-P	603 (314-845)	308 (-240-830)	462 (-1109-1237)
Vel	2,0 (-0.5-5.0)	2,0 (0.6-4.1)	-0,2 (-9.2-8.0)
Jerk	0,2 (-0.7-0.9)	-0,5 (-1.2-0.8)	0,4 (3.8-2.5)
#ROM	5,1 (2.8-7.5)	1,8 (-1.8-5.4)	2,5 (-6.8-11.8)
2D	246 (108-402)	24 (-255-334)	164 (-901-474)
3D	21 (9-31)	11 (-14-33)	22 (-70-51)

P-P =Phase-plot Area (degree²/sec), Vel = Maximum flexion velocity (degree/sec), Jerk = Jerk Index (number of changes in acceleration from full extension to full flexion), ROM= Range of motion (degree), 2D = 2-dimensional circumduction area (degree²), 3D = 3-dimensional circumduction area (cm²), RMDQ = Roland Morris Disability Questionnaire, BPI =back pain intensity, # Mean calculated instead of medians because ROM was normally distributed.

In general, low and non-statistically significant correlations were found ranging from no correlation to $r=-0.37$ between motion parameters and patient-rated outcomes. The 2D circumduction motion parameter was the only kinematic variable significantly correlated with pain score and RMDQ score in the total cohort ($p<.01$), but the correlation coefficients were low, ranging from $-.20$ to $-.22$ (Table 3A, Manuscript IV).

For sub-groups based on pain distribution, there were considerable differences identified in several motion parameters when comparing changes for participants with back pain only to those who had back pain and leg pain (Table 5, Manuscript IV). For patients with back pain only, all motion parameters changed in the hypothesized direction i.e. the motion parameter scores increased when clinically relevant changes were reported both relative to back pain intensity and to RMDQ score, with the exception of the Jerk Index. For the participants with back pain only and no clinically relevant improvement, zero was included in the 95% confidence intervals. This was different to those who had a clinically relevant improvement, with the exception of the Jerk Index and the median confidence intervals for the maximum flexion velocity parameter. Significant differences between the clinically relevant improvement and no clinical relevant change were found for the motion parameters ROM (5.9 (95% CI 2.3-9.4) vs. 1.0 (95% CI -3.7-5.7) degree) and phase-plot area (993 (95% CI 711-1182) vs. 161 (95% CI -558-828) degree²/sec) for back pain intensity in the back pain only subgroup. Significant increase in ROM (5.3 (95% CI 1.9-8.8) vs. -0.4 (95% CI -5.4-4.6) degree) was found between clinically relevant improvement and no clinical relevant change groups for back-related function (Table 5, Manuscript IV). This was in contrast to the group with back and leg pain, where several motion parameters changed in the opposite direction to that hypothesized and no significant differences was found. Overall, stronger correlations were found in the group with back pain only compared to the group with both back and leg pain, and the motion parameters showing the strongest correlations to patient-rated outcomes were phase-plot area, sagittal ROM and 2D circumduction motion (Table 3B, Manuscript IV).

Finally we assessed the relationship between treatment on pain and back-related function versus motion parameters (Table 6, Manuscript IV). The supervised exercise therapy group experienced the highest percentage of clinically relevant improvement compared to the other groups i.e. 85% for back pain and 77% for RMDQ. This was followed closely by the other two groups that ranged from 60% to 70%. In general, the exercise groups had higher regional lumbar motion change scores than the spinal manipulation group when no clinically relevant changes were reported (Table 6, Manuscript IV). Significant increase between the clinically relevant improvement and no clinical relevant change was found for the motion parameter phase-plot area in the SMT group (back pain 978 (95% CI 458-1309) vs. -305 (95% CI -880-858) degree²/sec) and (RMDQ 1054 (95% CI 467-1374) vs. 267 (95% CI -394-864) degree²/sec) (Table 6, Manuscript IV). The relative difference in the regional lumbar motion change scores between clinically relevant improved versus no clinical relevant change

was consistently higher for the SMT group (Table 6, Manuscript IV) and were stronger correlated than the exercise groups (Table 3 C, Manuscript IV).

Discussion

The first part of the discussion addresses each aim separately and our results will be compared to findings in the literature. The second part describes methodological considerations (strengths and weaknesses) for the systematic literature review and the cohort studies.

Discussion of main results

The quality of reporting in studies on reliability of 3D regional lumbar motion measurement systems

The findings from the systematic literature review indicate that acceptable reliability coefficients can be obtained by some 3D regional lumbar measurement instruments for a variety of motion parameters but, due to incomplete reporting, these estimates are difficult to interpret. Consequently, such measurements appear to be sufficiently reliable for assessment at a group level and for research purposes but not at an individual patient level in a clinical setting. Considerable variation was found in the quality and completeness of reporting between the domains.

Generally, the study populations evaluated appear to be relatively small sample sizes (ranging from 6 to 31) and in most studies, the populations do not represent a clinical population. By and large, colleagues or students were recruited to participate in the studies. A few studies did investigate subjects with LBP; however we found no difference in the level of reliability between subject groups with or without LBP [193;199;200]. When comparing reliability estimates between LBP and non-LBP patients, although the ICC may be relatively similar, the underlying variance components may be different. A likely scenario is that both the between-subject variance and the within-subject variance (measurement error) are higher for patients with LBP compared to those without LBP, but in the end provide similar ICCs [204].

All fundamental controllable procedures and factors that may influence the reliability should be reported. With appropriately detailed information on the test procedures, it should be possible for other researchers to replicate the methods on different samples or devices. This is also necessary if we are to enhance our understanding of mechanisms affecting measurement variability and thereby the clinical usefulness of these instruments. Shirley et al. have shown that, depending on the devices and procedures used, the lumbar sagittal ROM within a chronic LBP population varies significantly [205].

Several different methods of statistical analysis have been reported in the studies included in this review, many of which were reported in an incomplete manner and others were inappropriate for this type of research. As statistical analyses form the basis for drawing conclusions and comparisons between studies, there are obvious limitations to the usefulness of this body of literature. Furthermore, incomplete reporting about the normal

distribution and heteroscedasticity of the data is common and because errors often depend on the size of the measurement (larger error being associated with larger mean score) [206], sub-optimal statistical methods may have been applied.

To our knowledge, this is the first review summarizing the studies on reliability and evaluating the quality of reporting in this body of literature. The importance of reliability studies as a foundation for this field of science is gaining recognition; however overall, this review highlights the need for new high quality studies in order to identify and quantify factors affecting measurement error and reliability. Comprehensive reporting of all known and plausible factors affecting measurement error and thereby reliability is fundamental if future reliability studies are to effectively compare study populations, testing circumstances and reliability estimates. Generalizability studies could be a tool to study these sources of variance and investigate whether standardizing measurements has the potential to reduce measurement error.

The reliability and measurement errors of regional lumbar motion in chronic LBP patients

We found that the LOA were relatively wide and the reliability low to moderate depending on the particular motion parameters being investigated. Kinematic data from this group of chronic LBP patients may be sufficiently reliable as measurements of groups of patients, however because of the large LOA, this test procedure appears unusable at the individual patient level. Furthermore, reliability and measurement error vary substantially between subgroups of patients.

The reliability of these measures, predominantly ROM, has been investigated in a range of studies as reported in the systematic literature review [174] (Manuscript I). Any direct comparison between these studies and our study is difficult because of the variation in study populations, test procedures (examiner characteristics, measurement equipment, methods), choice of statistics applied and the level of detail reported. This may help to explain why we generally found lower ICCs and higher measurement error. The reasons for this discrepancy are probably many, however one fundamental reason is that our data were not initially designed to be used for test-retest analysis. Previous studies of reliability and measurement error in lumbar motion recordings have primarily used intra- and inter-examiner designs, mainly in small samples of healthy populations (Manuscript I). Our design did not provide the means to do a ‘true’ intra- and inter-tester design; however we did stratify relative to examiner (same or different examiner). As the same participants are not present in all of these stratified groups (same or different examiner), they cannot be directly compared. However, it does provide an indication of the amount of measurement error added when different examiners are used. Several other factors that might affect the reliability have been discussed previously in the method section. All motion parameters examined had smaller LOA for the ‘same examiner’ group, except for the Jerk Index (Table 5).

To help interpret the “ratio limits of agreement” (Table 4 and 5), if the patients estimated maximum flexion velocity is 20 °/sec on the first test then the second may differ from another by an estimate as low as $20 \text{ °/sec} \times 0.51 = 10.2 \text{ °/sec}$ or as high as $20 \text{ °/sec} \times 1.65 = 33 \text{ °/sec}$ on the next test (when using LOA ratio from pain stable subjects Table 5). For a subject with a higher velocity e.g. 60 the second weeks performance may be as low as $60 \text{ °/sec} \times 0.51 = 30.6 \text{ °/sec}$ or as high as $60 \text{ °/sec} \times 1.65 = 99 \text{ °/sec}$. Therefore these ratio limits vary in absolute terms but remains a constant ratio or percentage change from test 1 to test 2. Even though these ratio limits is wide they are more realistic in the way they are allowed to vary depending on the level of patients’ velocity [206].

In contrast to previous studies, we also examined the effect of other sources that might contribute to error i.e. the pattern of pain distribution, the influence of obesity, gender and pain level. The results for measurement error indicate that smaller LOA were found in the subgroups with patients with stable pain levels, patients with BMI below 30 kg/m², male patients, and patients in the Quebec Task Force classifications group 1 (back pain only). For the ICC_(1,1), the same pattern was present i.e. higher coefficients were generally found in the same subgroups; although this pattern was not as consistent as for the LOA.

The question still remains as to why these differences in reliability between subgroups seem to be present. The explanations may be related to biological as well as psychosocial factors. Following the rationale that pain influences motion through biological as well as psychosocial pathways [207], the difference between subgroups with stable and unstable pain makes sense. These findings indicate such a link because the unstable pain group generally exhibited a larger measurement error. With regard to BMI, the amount of fat tissue directly affects the distance between the instrument and the spine. A thick fat layer may induce increased measurement error through wobbling and sliding of the straps and/or through failing to palpate and fixate the instrument at exactly the same location on each test day. The tendency of larger measurement error in the subgroup with BMI above 30 kg/m² supports this “biological” theory; however psychosocial factors could also contribute. The difference between genders is surprising and the explanations may involve complex interactions of biological as well as psychosocial factors. Biological explanations might be based in female gender characteristics such as the distribution of body fat, potential discomfort due to pressure on the breasts from the equipment straps, or fluctuation in hormone levels. The response to psychosocial stress under biomechanical testing in a research environment has been found to be different between genders. Marras et al. found that women reduced their hip motion during the experimentally induced stress condition whereas men’s hip motion was relatively unchanged [207]. Phenomena like these might explain some of the difference in the measurement error. Likewise, some of the measurement error between same or different examiner(s) could be explained though psychosocial factors

although the majority of this variance probably is located in the practical aspects of the testing procedure e.g. fixation of the instrument and palpation of fixation points. The larger measurement error in the subgroup with back and leg pain could be explained by a more complicated biological disorder involving extra biological structures such as nerve roots, or referred pain from joints or muscles that are producing the leg pain. These extra structures may induce some extra noise. However psychosocial factors may also play an important role.

Regional lumbar motion change following treatment interventions

The assessment of changes in regional lumbar motion following intervention with either SET or SMT showed that regional motion changes can occur in chronic LBP patients over a 12-week period and that different treatments resulted in dissimilar changes. The group receiving spinal manipulation changed significantly in all, and the exercise groups in half, of the motion parameters included in the analysis. The spinal manipulation group changed to a smoother motion pattern (reduced Jerk Index) while the exercise groups did not (Figure 12 and Table 6).

Changes in motion properties of the lumbar region in response to manual therapy (mobilization or manipulation) and exercise have been investigated previously, but have shown conflicting results. In general, the literature is limited to studies measuring ROM, and therefore the following discussion focuses on these measures. Following flexion mobilization in individuals with LBP, Konstantinou et al. found statistically significant, but small, immediate increases in mean spinal ROM, compared with the placebo [208]. In contrast, Goodsell et al. found no change in the mechanical behavior of the lumbar spine of patients with LBP following posterior-anterior mobilization [209]. Similarly, Petty found no change in ROM following 2 minutes of posteroanterior mobilization on 14 asymptomatic volunteers [202]. Moutzouri et al. found that Mulligan's Sustained Natural Apophyseal Glide mobilization did not demonstrate significant differences in flexion ROM when compared to sham mobilization on a sample of 49 asymptomatic volunteers [159].

For manipulative therapy, Burton et al. found a significant increase in mean mobility in the first month [210]. Recently, Stamos-Papastamos et al. published a study using same-subject, repeated measures, crossover design examining manipulation and mobilization on 32 asymptomatic subjects and found no significant effect on flexion and extension ROM; but some individual variations in effect were observed [154]. Similarly, Lehman and McGill did not find consistent short-term effects of manipulation on ROM in a sample of 14 non-specific LBP patients [211] and in a recent systematic literature review Millan et al. concluded that none of the retained studies showed an immediate effect of SMT on lumbar ROM [212]. In our study, we did find a statistically significant difference in sagittal ROM following manipulation treatment (0.011), although the pre-to-post group difference of 4.4 degrees was small (Table 6).

For exercise therapy, Tamimela and Härkäpää examined 143 chronic LBP patients attending a 12-week multidimensional back treatment program that included different exercises and found increases in lumbar spinal ROM [213]. Magnusson et al. found that functional rehabilitation increased motion and velocity in 27 patients with chronic LBP measured by a triaxial goniometer [57]. These findings differed to some extent from our results, although they both used a study population relatively similar to ours. In our study, the mean ROM and velocity did increase slightly for both the HEA and SET groups, but only ROM for the HEA group increased significantly ($p=0.005$). The explanation for these divergent findings may be differences in the exercise interventions, measurement technologies or other factors.

SMT treatment aims directly at restoring function in joints that have a physiological dysfunction. This is done by directly applying forces (SMT) on the spinal structures that restore function by different pathways e.g. direct structural mechanical changes in the spinal joints and/or neuromuscular pathways as described in the introduction section. This study supports these theories because significant changes were found in all motion parameters. The pathway through which exercise may modulate spinal mobility may be different i.e. it seems reasonable to believe that it could be a more general pathway because the exposure is unspecific and the neuromuscular system is activated through active pathways. The different way that these treatment modalities involve the neuromuscular system may partly explain the difference seen in the Jerk Index (smoothness of motion).

Relationships between changes in regional lumbar motion and changes in patient-rated outcomes

In daily clinical practice, many decisions are at least partly made on the basis of more or less objective measures. For instance, if a clinician finds a patient has reduced lumbar motion, he/she could prescribe stretching exercises or perform manipulative treatment in order to improve mobility and ‘theoretically’ when movement is restored, the patient will also feel less pain. The idea that the constructs of ROM and LBP correlate may be based on the widespread belief that back pain is caused by biomechanical factors that manifest as ‘inappropriate’ motion patterns or restrictions and by ‘correcting’ the fundamental biomechanical problem, this will then reduce the pain level in a reasonably predictable way [214;215]. However, our results could not confirm that such a link exists.

The relationship between change scores in regional lumbar motion and patient-rated outcomes such as pain and RMDQ in chronic LBP patients treated with either SET, HEA or SMT was found to be weak or non-existent. Therefore, this study provides evidence that regional lumbar motion changes do not appear to be valid measures of chronic LBP patients’ perception of improvement in their pain intensity score from the past week or back-related function and therefore should be interpreted and used cautiously in the clinical setting.

However, regional lumbar motion and patient-rated responses to treatment differed between patients depending on the presence of leg pain. Our results indicate that the kinematic response to treatment is less predictable when

leg pain is also present (Table 3B and 5, Manuscript IV) and these findings could be related to the underlying biology causing the pain. In addition, these results indicated that patients receiving 'active' treatment in the form of exercises showed regional lumbar changes regardless of their perception of improvement (pain level and patient-rated function), whereas in patients receiving the more 'passive' SMT, patients' perception and regional lumbar changes were more consistently associated. These findings may be related to the neurobiological mechanisms related to the different treatment modalities.

Relationships between lumbar motion and patient-rated outcomes have been examined in several studies in the literature, predominantly using ROM measurements. Zuberbier et al. reviewed the literature on convergent and discriminant validity of lumbar ROM tests for the characterization of LBP and injury [165]. Convergent validity (whether the test scores are associated meaningfully with the parameters to which they should be related: in this case correlation with spinal disability and self-reported function [165]) was reported to be inconclusive [165]. However, most commonly, the coefficients of correlation were below 0.5, indicating that within a group of patients, the variance in ROM measures accounted for a maximum of 25% of the variance in self-rated disability. Studies not included in this review also indicate varying degrees of correlation. Cox et al. found a correlation between unloaded ROM or velocity versus pain or disability of 0.41-0.55 [216] and similarly, Kang et al. have shown that an individual's range of lumbar motion is significantly related to his/her subjective pain ratings, although the relationship is weak and only explains a small amount of the variance [217].

Using the CA 6000, McGregor et al. found that the strength of the relationships between subjective clinical findings and objective clinical tests in 138 LBP patients was not as large as anticipated i.e. the percentage of explained variance was low (r^2 between 0.16 and 0.45)[218]. Parks et al. used a 3D lumbar motion instrument to examine the correlation between simple lumbar end ROM measures and functional ability, which they found to be weak or non-existent [219].

In our study, we have not reported correlations at the cross-sectional level as reported in the above-mentioned studies. However, following an explorative (Spearman) correlation analysis on the cross-sectional level at baseline (two) between all motion parameters and back pain intensity (0-10) or RMDQ score, we found correlations of 0.0 to -0.09 relative to pain (none being significant) and 0.0 to -0.23 compared to RMDQ (four parameters being significant $p=0.05$) indicating a non-existent or weak correlation at the cross-sectional level. The 2D circumduction parameter was the one most correlated to RMDQ.

Relatively few studies have examined changes in lumbar motion and their relationship to pain or function in LBP patients receiving treatment (i.e. a longitudinal perspective). We did this in our study (Manuscript IV). Burton et al. studied lumbar sagittal mobility and low back symptoms in 55 patients (pain duration of more than a month in 64% of the patients) treated with manipulative therapy. They concluded that if any benefits actually

result from manipulative therapy, they are not a direct function of overall lumbar sagittal mobility because symptomatic improvement was as common in patients with unaltered or reduced mobility as it was in those who showed an increase in mobility (ROM). Furthermore, they concluded that changes in mobility at one month had no value for predicting symptomatic status at one month or one year [210]. Taimela and Harkapaa examined the association between pain reduction and objective measurements for improvement in physical functioning in 143 chronic LBP patients following a 12-week multidimensional back treatment program. Their results indicate that pain reduction might be associated with improvement in spinal mobility but the correlations between physical functioning parameters (mobility and strength) and pain reduction were low and not associated with the magnitude of the improvement [213]. Similarly, Poitras et al. examined work-related back pain in 111 patients with sub-acute and chronic back pain randomized to four different methods of management and concluded that lumbar spine kinematics, ROM and spinal velocity during flexion and extension of the trunk did not appear to be valid measures of disability in patients with sub-acute and chronic back pain [161]. Our study is therefore generally concordant with the literature, i.e. that association between lumbar motion measures is weakly related to patient-rated outcomes in chronic LBP patients.

McGregor and Hughes examined the ability of motion analysis to detect long- and short-term changes in performance following lumbar decompression surgery on 52 patients [145]. Improvements in back pain, leg pain and the Oswestry Disability Score were reported; however, change in motion was not correlated with change in functional disability scales or pain, thus further supporting the view that objective and subjective measurement tools are not linearly related to each other.

For discriminant validity (described by Zuberbier et al. to indicate whether the test scores are able to distinguish between different populations that would be expected to show different degrees of lumbar spine mobility) ROM measures seem to have little capacity to distinguish ‘healthy’ from LBP patients [165]. Studies using more advanced techniques such as neural network and advanced classification analysis have been shown to be better for distinguishing between LBP patients and healthy individuals [54;57;58;60;220;221] .

One of these studies by Dickey et al. investigated the relationship between vertebral motion of the lumbar spine and associated pain in a select group of nine chronic LBP patients using a percutaneous method. Using a neural network model, they found a strong relationship between observed and predicted pain ($r^2 = 0.997$), however the nature of these relationships was nonlinear. Linear correlation and linear discriminant analysis did not effectively describe these relationships, i.e. the highest reported correlation between a ‘single’ motion parameter and the reported pain was -0.467, ($r^2 = 0.218$) [221]. Cherniack et al. showed that the Lumbar Motion Monitor instrument and quantification method as described by Marras [60] appears to be a useful assessment tool for

gauging the presence of LBP and low back disorder [220]. However we did not evaluate healthy people in our study and consequently did not evaluate discriminative validity.

In summary, the literature and our own research, indicates a weak association on a cross-sectional level as well as from a longitudinal perspective between patient-rated outcomes and different lumbar motion measures. Explanations for this lack of association may be many. Overall, the study samples may have been too heterogeneous and therefore the participants responded very differently to treatment. Some of the ‘signal’ (movement) may have disappeared in ‘noise’ (random or systematic measurement error), introducing limited usability. In addition, changes in physical (e.g. ROM) and mental (e.g. pain) outcomes may not be related at the same time point, i.e. it may take a variable amount of time before ‘objectively improved function’ translates into patient-perceived improvement, especially in chronic patients, where depression, etc. may complicate and delay the recovery process. Or, it may just be that the hypothesized mechanism that forms the foundation for some treatment modalities and clinical decision-making is not as we thought it was, at least for chronic LBP patients. One explanation for the lack of correlation between lumbar motion and back pain may be that the psychosocial factors have a higher impact on the condition than biological factors. When the condition has become chronic, the complexity increases and in some cases the biological factors that caused the pathology may no longer be present, but the pain remains.

Methodological considerations

Systematic literature review

This research commenced with a systematic literature review, which is considered an excellent tool for gaining a broader view of a complicated topic. In our particular case, we were unable to find any generally accepted quality checklist for reviewing reproducibility studies, and therefore, we designed our own. The checklist we developed was based on a broad representation of relevant published material including reviews and guidelines for reporting reliability and agreement studies (GRRAS)[169;177-179], the standards for reporting of diagnostic accuracy (STARD)[180;181]and quality assessment of diagnostic accuracy studies (QUADAS) [182-184]. In addition, we designed the checklist specifically to guide the systematic extraction of information in a standardized way [175;176]. The checklist was completed for each included manuscript by two of the authors independently, the lists were then compared and disagreements resolved by discussion. The use of alternative quality criteria might have resulted in a different quality appraisal, although it may not have changed our interpretation of the results. Furthermore, there may be relevant studies published in other languages, which were not included. We did a comprehensive computerized search in collaboration with a health science research librarian using a carefully developed search strategy and electronic searches were performed in Pubmed, CINAHL, Embase and Mantis databases. Although the total number of citations identified in the electronic databases was relatively high (2,042), we did find and include four more after careful scrutiny of reference lists. One challenge was that often our topic of interest was hidden as a secondary aim, and therefore could not be easily identified.

A number of reporting biases could have affected our results such as publication bias, language bias, and citation bias. We weighted the perceived chances of finding relevant papers against the time resources available and decided not to include the grey literature i.e. unpublished studies (potential ‘publication bias’) and not to perform a hand-search of selected journals in addition to the primary search. By careful scrutiny of reference lists, we found an additional four papers for inclusion, which could give rise to ‘citation bias’. These biases and limitations may have affected our results.

Cohort studies

In the cohort studies, there are strengths and weaknesses that need to be highlighted related to the design, study sample, lumbar motion test procedure and management of motion data.

Design

The cohort used for three manuscripts in this thesis was a subset of the study population from an observer-blinded, parallel-group, randomized clinical trial by Bronfort et al. [1], on which we performed a secondary analysis. The strengths and weaknesses related to this original RCT are therefore also relevant for our studies.

The strengths of the primary study include the strict adherence to concealment of randomization, blinding of objective examiners, and an intention-to-treat analysis. Furthermore, the number of participants was appropriate to achieve adequate power and the compliance with treatment was high. However for the current studies presented in this thesis, one limitation is that they were not specifically designed and powered for the motion analysis evaluation, i.e. these were secondary measures of effectiveness. Therefore, the reliability study may seem to have an unnecessarily high number of participants compared to studies presented in the literature [174]. On the other hand the high number of participants offered the opportunity to stratify into subgroups, and provide insight into factors affecting measurement error and reliability. The intervention studies presented in this thesis also have the limitation that the number of participants was not powered specifically to the regional lumbar motion assessments but to the effect size difference between groups in both the short and long term [1].

A large battery of assessments including patient-rated outcomes, trunk performance measures and qualitative measures was performed, forming a rare and comprehensive body of material. This may however have produced some negative consequences such as patients losing concentration and precision when answering questionnaires or performing tests simply due to the large volume of tasks involved in the test battery.

The study was not developed to differentiate between the specific effects of treatment and contextual effects (including patient/provider interactions, attention, etc.) that could explain some of the patient outcomes. Instead, the trial was designed to be pragmatic in nature, investigating typical interventions offered in clinical practice. Therefore, the HEA group was not a stringent control group but instead was intentionally minimal in its approach so it could serve as a pragmatic control. Other limitations of the study design include the potential impact of the loss to follow-up and missing data, as well as the lack of blinding of patients and providers, as these ideal attributes are not feasible in exercise trials.

Patients

Study participants were 18 to 65 years of age with mechanical LBP of at least 6 weeks' duration. Most patients had experienced pain for more than a year (~80%), however the duration varied. This may have reduced the homogeneity of both their mental and physiological characteristics, thereby influencing their potential to respond to these treatments. In addition, for this study group with mechanical LBP, the underlying pathoanatomical etiology of the pain probably varied and the response to treatment was therefore also likely be diverse.

Participants were recruited through local newspaper advertisements, community posters, and postcard mailings, with the consequent limitation of only representing people exposed to these media who had the ability to read them. This may have biased the sample by having an overrepresentation of people with a higher education level, which has been reported to be associated with an increased ability to cope [222].

Exclusion criteria were implemented in order to achieve a relatively homogeneous and stable cohort with consistent symptoms and severity of LBP, where the level of pain intensity (ordinal 11-box scale ≥ 3) made it possible to measure changes over time and where other health conditions were unlikely to influence the outcomes during the one year clinical follow-up period. At the group level, pain intensity was relatively stable (Table 3), however individual fluctuations in pain levels were observed.

Patient characteristics are reported in Table 3 and in further detail in the primary study [1]. One specific aspect of the participant characteristics that could have influenced the kinematic measurements was the relatively high BMI (mean $>28 \text{ kg/m}^2$) for both genders. The WHO classifies a BMI of $18.5\text{-}24.9 \text{ kg/m}^2$ as normal, a BMI of $25\text{-}29.9 \text{ kg/m}^2$ as overweight or pre-obese, and a BMI of 30 kg/m^2 or more as obese. In addition, increased BMI is associated with increased morbidity and mortality [223].

As mentioned in the primary paper by Bronfort et al. [1], the patients in all three groups had greater expectations of improvement for SET than for HEA. This was likely to be due to obvious differences in dose and attention and this may have biased the outcomes following the interventions. Finally, the individuals examined in the current study were chronic LBP patients and therefore, the results should not be generalized to other populations.

No significant difference between the three treatment groups SET, SMT and HEA in terms of pain and other patient-rated outcomes, in short- and long-term were found in the primary analysis [1]. Based on these results we found it acceptable to collapse these treatment groups in order to analyses associations in changes in pain and back related function versus regional lumbar motion in the total cohort (Manuscript IV). However it is a bias that may have affected our results.

Lumbar motion measurement test procedures

The objective assessments were repeated on two baseline test days for all participants, which might have helped to reduce the bias caused by the learning effect, as the second baseline recordings are used for comparison with the post intervention recordings. The lumbar dynamic motion performance measures were performed by blinded examiners, who were trained to the standard deemed necessary by the lead investigators. The use of highly skilled personnel may have strengthened the outcome precision, but on the other hand, may limit the applicability to clinical practice.

A stringent protocol was followed by the examiners including detailed instructions on equipment attachment procedures and the conduct of motion trials. A general concern when performing such assessments in a research environment is that we may not be measuring what is 'normal' for the patient. Research using sophisticated biomechanical equipment inherently creates an artificial situation that may influence the usefulness of the outcomes.

A limitation of the current study protocol may be that the time of day was not controlled for. Research has indicated that ROM measurements are influenced by the time of day [224]. Ensink et al. showed that the total lumbar ROM measured by an inclinometer technique and the modified-Schober sign increased significantly throughout the day from morning to afternoon. It is therefore likely that some amount of measurement error may have occurred in our recordings due to this factor.

For the current study, a range of other aspects may potentially have affected the amount of measurement error or 'noise'. One of these is the speed at which the regional lumbar motion tests were performed. This has been evaluated by McGregor and Hughes using the same equipment as we used during a flexion/extension test at three different speeds: slowly, preferred, and as fast as they could [200]. In addition, this was evaluated both on subjects without back pain and subjects with central spinal stenosis. The results indicated that testing at a participant's preferred speed produced more consistent readings of motion characteristics (in terms of both ROM and velocity), which supports the method used in the current study. Another factor, well-known for introducing measurement error (random and systematic), is when different raters perform an assessment of the same patient (inter-rater reliability). Nine different examiners performed the lumbar dynamic performance test and a weakness with the current study is that for the second baseline and the post intervention assessments, only 36% were performed by the same examiner on each test day. Other factors affecting measurement error are the movement of the equipment on the skin surface. In the current study, this is probably exacerbated by the patient's wearing a t-shirt (resulting in a layer of fabric between the skin surface and the manufacturer-supplied belt).

"Securing a measuring device on a subject is accepted as problematic and not unique to this instrument or to measurements of the spine in particular." [225]

Troke et al. found the need to improve the CA 6000 measurement equipment because difficulties were encountered with the manufacturer's strap system. They developed a modified fixation method using double-sided tape (3M Co) and judged it to be accurate, practical and ethically acceptable[225]. The use of the manufacturer-provided strap system may therefore be considered a limitation by some researchers.

The present setup, where the fixed extremity of the device was mounted on the sacral crest (S2) and the mobile end of the device was mounted at the level of T7 with the top edge of the horizontal metal pieces aligned evenly with the inferior angles of the scapulae, may result in some measurement error caused by a lack of precision in the placement of the equipment. However other methods e.g. counting spinous processes from the sacrum may also introduce bias, especially in obese people, and this procedure is more time-consuming. In obese people, the inferior angles of the scapulae are easier to localize than palpation and counting each spinous process. Furthermore, the results of our studies are also influenced by the lowest thoracic region being included in the motion analysis.

Management of motion data

The program ‘spinal motion analyzer 1.0’ was developed in collaboration with a skilled programmer in order to examine the comprehensive and complex 3D regional lumbar motion data. To reduce the complexity and volume of data, the recorded 3D regional lumbar motion data were reduced to a number of separate motion parameters. The selected parameters could therefore be considered as reductionist models that describe complex regional lumbar movement patterns in LBP patients. The motion parameters were developed from ideas and hypotheses generated from clinical experience and expert opinions from experienced researchers. More than 100 motion parameters were developed and extracted by means of the computer program. A few parameters were selected for further statistical analysis; including the most commonly reported parameters such as ROM and velocity parameters, and more complex parameters based on time-frequency analysis and area calculations of the trajectory movement of the spine [57;161;226]. Obviously, the choice of parameters and analysis means that other types of important kinematic information may remain unexplored. Other strategies such as data mining may have been useful, however this is subject to debate.

A further limitation of the results is that they cannot be extrapolated to other technologies. Assessments of agreement between measurements of the OSI CA 6000 and other technologies (Fastrak) have indicated that the two devices do not always yield comparable measures for lumbar mobility but each can be used reliably in longitudinal studies. Comparison of values for lumbar mobility must be considered device-specific [173].

Some specific strengths and limitations must be mentioned with respect to the selection process used in the clinical trial. For this process, a computerized comparison with an additional manual visual inspection of all sagittal plane plots for each patient on each test-day was performed using the developed program. The inspection process was limited by the inherent subjective judgment but it was also a clear strength in that it made it possible to identify potential errors not detected by the computerized comparison. Any subjective judgment or manual override was however rarely needed (less than 10 occasions). Therefore, the purposes of this program were to

choose and extract the estimated two best data trials, and in addition, to detect missing or erroneous data. The selection process was based on a computerized comparison of all movement trials in each set in relation to ROM, uniformity of curves, and length of curve, and a manual visual inspection of sagittal motion plots.

Examples of the instrument defects detected during the process are presented in Figures 13 and 14. During the data collection period, a technical defect occurred in one of the three high precision potentiometers unfortunately precluding the determination of sagittal plane motion. Before the instrument defect was discovered by the clinicians and the CA 6000 was sent back to the factory for correction, some regional lumbar motion recordings had been carried through. All trial were examined closely by visually plots of all trails in all motion directions using the program 'spinal motion analyzer 1.0'. In addition, all outliers ($>2SD$) for all motion parameters were examined again; however no extra errors in any motion direction were found. This error turned out to involve six patients who therefore could not be used for analysis.

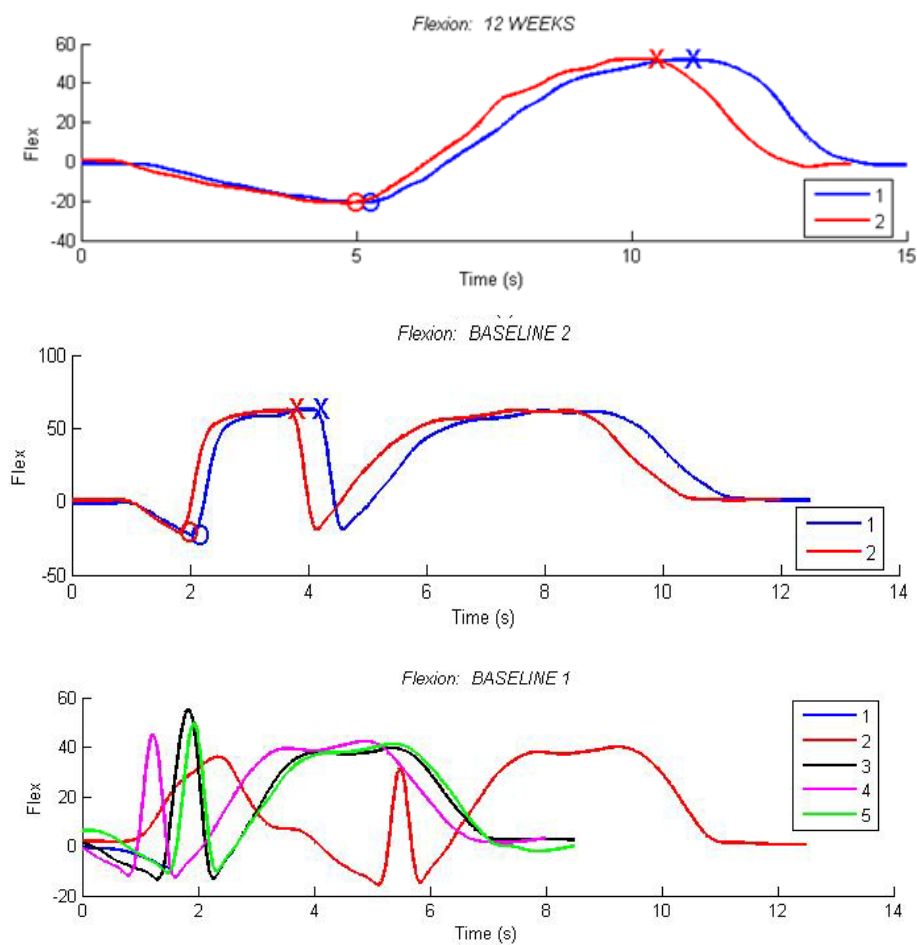


Figure 13: The uppermost graph illustrates an example of a normal extension/flexion motion trial. The two lower graphs show instrument errors in the sagittal (extension/flexion) test procedure. During the first part of the movement (approximately full extension) a sudden and very quick unintended shift in positive direction occurs at a rate probably not humanly possible. In these cases, the data could not be used because all trials were affected with instrument error and therefore the patients were excluded from the kinematic analysis.

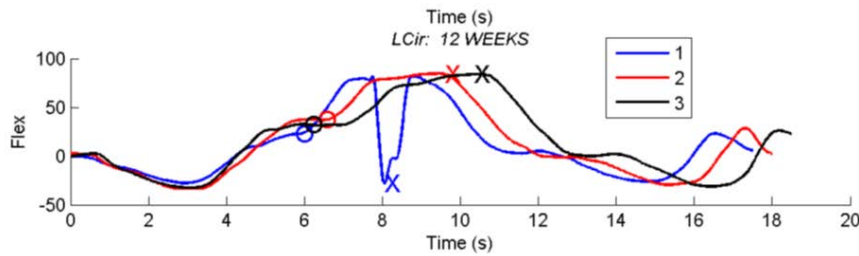


Figure 14: Sagittal plane movement data from three trials of a circumduction test procedure. Trial one (blue line) is corrupted by an instrument error and therefore discarded from the analysis whereas the two subsequent trials (trials two and three) are accepted for further analysis.

With regard to some of the motion parameters, some specific limitations must be emphasized. The phase-plot area was calculated based on cross-product calculations between vectors drawn from the neutral position to each coordinate point in the phase-plot (Figure 7). As the initiation and termination of the movement may be difficult to define, we chose to calculate the phase-plot area during the entire sampling period of the trial. As a consequence, accumulation of ‘noise’ obtained from small movements occurring before and after the actual movement may have influenced the results. Nevertheless, as this ‘noise’ occurred close to the neutral position (0, 0), hence little accumulation, the effect is likely to be negligible.

The Jerk Index was calculated from the maximum extension to the maximum flexion position as the mean frequency of the first derivative of the acceleration multiplied by the movement duration i.e. the number of changes in acceleration during movement from maximum extension to maximum flexion. Therefore, the present Jerk Index is very dependent on motion time. Accordingly, two movements performed with different velocities may be performed with similar number of jerk oscillations per second but result in different Jerk Index, because of differences in movement time. There may be several other ways to calculate this motion parameter e.g. to calculate the total number of jerk oscillations divided by the movement distance.

Conclusions

The main conclusions of this Ph.D. thesis are:

- 1) Many different instruments used for the evaluation of regional lumbar motion have acceptable reliability coefficients. Incomplete reporting of study samples, test circumstances, testing protocol and statistics makes these reliability estimates difficult to interpret. Most instruments used under standardized conditions may only be considered sufficiently reliable to be used for research purposes at the group, but not at the individual, level.
- 2) The lumbar regional motion parameters displayed by chronic LBP patients using this standardized protocol and instrumentation can be used for group comparison but not to assess individual patients because of their commonly large measurement error. Reliability and measurement error of regional motion parameters vary between sub-groups of patients with chronic LBP. The reliability estimates for the pain stable subgroup ranged from 0.55 to 0.79 which is close to the hypothesized range.
- 3) The group receiving spinal manipulation changed significantly in all, and the exercise groups in half, of the motion parameters included in the analysis. The spinal manipulation group changed to a smoother motion pattern (reduced Jerk Index), while the exercise groups did not. Thus the hypothesis that both supervised exercise therapy and spinal manipulative therapy would change significantly in all motion parameters and the minimal intervention group would not change in motion parameters was only partially confirmed i.e. only confirmed for spinal manipulative therapy group.
- 4) Overall, changes in regional lumbar motion were poorly associated with back pain intensity scores measured by ordinal 11-box scale for the previous week and back-related function measured by RMDQ. No significant difference in regional lumbar motion parameters was found between clinically relevant improvement and no change group for the cohort as a whole. However, associations between regional lumbar motion versus patient-rated pain and back-related function were different relative to subgroups. Thus stronger correlation coefficients and significant differences between clinically relevant improved versus no clinical relevant change were found in some motion parameters in the subgroup with back pain only and the treatment group receiving spinal manipulation. Consequently the hypothesis that patients who had a clinically relevant improvement in pain and function would have greater change scores in velocity, ROM, circumduction area, and have a smoother motion compared to patients who did not achieve a clinically relevant improvement, could not be confirmed in general but the explorative analysis indicate that this could be different in certain subgroups.

Perspectives

Clinical implications

Early last century, Mencken stated *'There is always an easy solution to every human problem – neat, plausible, and wrong'*. [227]

The results from this work indicate that these regional lumbar motion outcome measurements of chronic LBP patients should be interpreted with caution because they appear to be neither ideal for evaluating improvements nor for determining severity of condition. However, simple objective kinematic assessments may still play a role in the clinical setting, although not as prominently as advocated by some. In addition, the story might be different for certain sub-groups or other spinal regions.

In clinics, measurement error is very important because patients are managed at the individual level. Measurement error cannot be reduced by the sample size as in research, only by improving measurement procedures or repetition of measurements which is often too time-consuming in daily clinical practice. Therefore, clinicians are advised to have a continuous focus on standardizing objective measurement procedures as much as possible in order to reduce measurement error.

This thesis provides evidence that different treatments like exercise and manipulation result in different kinematic outcomes. This is a small step towards explaining how different treatments affect regional lumbar motion.

Carey and Mielenz asked the question: *'If a patient is functioning well with limited ROM, should the patient's outcome be considered good or poor? Was the treatment a success or a failure?'* [228]. We think that in the end, it is the patient's perception that counts the most. Overall, this research indicates that objective regional lumbar motion measurements do not correlate well with patient perceptions and therefore should not be used as primary outcome measures in the clinic to determine treatment success for chronic LBP patients.

Research implications

Comprehensive and transparent reporting of all known and plausible factors potentially affecting measurement error and reliability are fundamental in future test-retest studies.

Design: Studies using generalizability theory would potentially be useful to study sources of measurement error and illuminate where aspects of measurement error are located. Key areas are likely to include the standardization of measurement procedures and instructions to examiners.

Patients: As regional lumbar motion measurements may not be relevant in the evaluation of all back pain patients, future research should focus on identifying subgroups where this technology could be relevant. For instance, we identified a sub-group effect using a rough stratification based on pain distribution and BMI. Furthermore, the relevance of kinematic assessment of acute back pain patients should also be evaluated.

Equipment and testing circumstances: The equipment used for kinematic assessment is generally very precise - the challenges appear when humans get involved. Therefore, procedures and attachment techniques should be further developed in order to reduce measurement bias. Future investigations should perform test-retest studies on all kinds of motion parameters, not only ROM.

New technology: In the near future it may be possible to measure patient kinematics throughout the day while at the same time registering pain. Thus registering and analyzing patterns of motion under ‘normal’ activities may provide insight into what causes pain, and in addition may allow the instrument to be used as a treatment tool. For example, it could alert the patient when doing pain-provoking activities and postures for longer periods.

Data analysis: Future studies could use other analysis techniques, such as data mining, in order to analyze the complex motion data from many different dimensions or angles, categorize them, and summarize the relationships. In our case, we have more than 100 motion parameters and a battery of patient-rated outcomes ready for this kind of analysis.

“Absolute clarity is the privilege of fools and fanatics”[229]

Summary in English

There are many theories concerning the etiology of low back pain (LBP) but in spite of considerable scientific effort, the definitive pathoanatomical and psychosocial pathways to LBP remain largely unknown. One way to investigate underlying biology and possibly sub-grouping of patients with LBP is by assessing regional motion and how this may vary between patients and possibly change over time as symptoms vary.

The overall aim of this work is to obtain a deeper understanding of the reliability of measurements for regional lumbar motion, to examine motion changes over time and their relationships with changes in pain and back-related function.

We conducted a systematic review of the literature dealing with reliability and/or measurement error of 3D regional lumbar motion measurement systems. Subsequently regional lumbar motion data from a subset of participants from a randomized clinical trial were used for reliability and longitudinal cohort analyses. Participants were 18-65 years of age with a primary complaint of LBP of at least 6 weeks' duration with or without radiating pain to the lower extremity that had no specific identifiable etiology but could be reproduced by back movements or provocation tests.

The systematic literature review (Manuscript I) broadly showed that the level of reporting was incomplete in several domains, i.e. study population, test circumstances, and data analysis and presentation, downgrading the quality of reporting in general and resulting in the reliability and measurement error estimates being difficult to interpret. However, acceptable Intraclass Correlation Coefficients (ICC) were found indicating that such instruments may be used for research purposes.

In Manuscript II, dealing with reliability of the regional lumbar motion measurements in our own data, we found generally lower ICCs and higher measurement errors than reported in the literature. We investigated variation in reliability between subgroups of patients and found that both reliability and measurement error varied between subgroups.

In Manuscript III, we investigated if treatments actually change regional lumbar motion by modulating regional lumbar motion, and whether specific treatment modalities affect regional lumbar motion differently. The group receiving spinal manipulation changed significantly in all, and the exercise groups in half, of the motion parameters included in the analysis. The spinal manipulation group changed to a smoother motion pattern (reduced Jerk Index) while the exercise groups did not.

In Manuscript IV, we found that the relationship between change scores in regional lumbar motion and patient-rated outcomes (pain-related disability measured with the Roland Morris Disability Questionnaire and pain measured with ordinal 11-box scale) were generally weak. However, associations between regional lumbar motion versus patient-rated pain and back-related function were different relative to subgroups. Thus stronger correlation coefficients and significant differences between clinically relevant improved versus no clinical relevant change were found in some motion parameters in the subgroup with back pain only and the treatment group receiving spinal manipulation.

Summary in Danish

Der er mange teorier om ætiologien bag lænderygsmerter (LRS), men på trods af en betydelig videnskabelig indsats, er de patoanatomisk og psykosociale årsager til LRS stort set ukendte. En måde at underindddele patienter med LRS og undersøge den bagvedliggende biologi er at måle lænderyggens regionale bevægelse og registrere om denne varierer mellem patienterne og ændres over tid i takt med at symptomerne ændres.

Det overordnede formål med denne afhandling var at opnå dybere indsigt i pålideligheden af målinger af den regionale lumbale bevægelighed, samt at undersøge, bevægelsesændringer over tid, og sammenhænge i patienternes smerter og ryg relateret funktionsnedsættelse.

Vi gennemgik den aktuelle litteratur omhandlende pålidelighed og/eller målefejl ved regional lumbal 3D måleudstyr systematisk. Desuden foretog vi reliabilitets- og longitudinelle kohorte analyser af bevægelsesdata fra en gruppe deltagere i et randomiseret klinisk forsøg. Deltagerne var 18-65 år, og havde LRS med eller uden udstrålende smerter til benene i mindst 6 uger som deres primære klage. Smerterne kunne reproducere ved rygbevægelser eller tests, men uden nogen specifik identificeret tilgrundliggende ætiologi.

Overordnet viste den systematiske litteraturgennemgang (Manuskript I), at rapporteringen var ufuldstændig i flere domæner fx studiepopulation, test omstændigheder, samt dataanalyse og præsentation, hvilket reducerer kvaliteten af rapporteringen i almindelighed og vanskeliggør fortolkningen af estimer for pålidelighed og målefejl. Vi fandt acceptable Intraclass Correlation Coefficienter (ICC), hvilket tyder på, at disse måleinstrumenter kan anvendes til forskningsformål.

I Manuskript II analyserede vi pålideligheden af de regionale lænderygs bevægelsesmålinger fra kohorten af kroniske LRS patienter og fandt generelt lavere ICC værdier og større målefejl end rapporteret i litteraturen. Vi undersøgte desuden variationen af pålidelighed mellem undergrupper af patienter, og fandt at både pålidelighed og målefejl varierede mellem disse undergrupper.

I Manuskript III, undersøgte vi, om behandlingerne ændrede bevægeligheden i lænderyggen, og om specifikke behandlingsmodaliteter påvirker ryggens bevægelse forskelligt. Den gruppe, der fik manipulationsbehandling af lænderyggen havde signifikante ændringer på alle bevægelsesparametre. Grupperne der modtog træningsterapi opnåede signifikante ændringer i halvdelen af de inkluderede bevægelsesparametre. Desuden opnåede gruppen der modtog manipulationsbehandling et mere glat bevægelsesmønster (reduceret Jerk Index) i modsætning til træningsterapi grupperne, der ikke opnåede et glattere bevægelsesmønster.

I manuskript IV undersøgte vi forholdet mellem den regionale lænderygs bevægelighed og selvrapporterede effektmål (smerte-relaterede invaliditet målt med Roland Morris Disability Questionnaire og smerte målt med ordinal 11-box skala) i gruppen af patienter med kroniske LRS, som blev behandlet enten med træningsterapi eller manipulationsbehandling. Vi fandt overordnet, at forholdet mellem ændringer i den regionale lænderygs bevægelighed og patienternes selv vurderede score, var meget svag. Men sammenhængen mellem den regionale lumbale bevægelse imod patient-vurderet smerter og ryg-relaterede funktionsniveau var forskellig i forhold til undergrupper. Således fandt vi stærkere korrelationskoefficienter og signifikante forskelle mellem klinisk relevant forbedrede imod ingen klinisk relevant ændring i nogle bevægelses parametre, i undergrupperne med rygsmerter uden udstrålede bensmerter og behandlings gruppen der modtog manipulations behandling.

References

- [1] Bronfort G, Maiers MJ, Evans RL et al. Supervised exercise, spinal manipulation, and home exercise for chronic low back pain: a randomized clinical trial. *Spine J* 2011;11:585-98.
- [2] Savigny P, Watson P, Underwood M. Guidelines Early management of persistent non-specific low back pain: summary of NICE guidance. *Brit Med J* 2009;338.
- [3] Deyo RA, Rainville J, Kent DL. What can the history and physical examination tell us about low back pain? *JAMA* 1992;268:760-5.
- [4] Hancock MJ, Maher CG, Laslett M, Hay E, Koes B. Discussion paper: what happened to the 'bio' in the bio-psycho-social model of low back pain? *Eur Spine J* 2011;20:2105-10.
- [5] Kent P, Keating J. Do primary-care clinicians think that nonspecific low back pain is one condition? *Spine* 2004;29:1022-31.
- [6] Kent P, Mjosund HL, Petersen DH. Does targeting manual therapy and/or exercise improve patient outcomes in nonspecific low back pain? A systematic review. *BMC Med* 2010;8:22.
- [7] Kent PM, Keating JL, Taylor NF. Primary care clinicians use variable methods to assess acute nonspecific low back pain and usually focus on impairments. *Man Ther* 2009;14:88-100.
- [8] Dagenais S, Tricco AC, Haldeman S. Synthesis of recommendations for the assessment and management of low back pain from recent clinical practice guidelines. *Spine J* 2010;10:514-29.
- [9] Andersson GBJ. Epidemiological features of chronic low-back pain. *Lancet* 1999;354:581-5.
- [10] Ladeira CE. Evidence based practice guidelines for management of low back pain: physical therapy implications. *Rev Bras Fisioter* 2011;15:190-9.
- [11] Underwood M, Watson P, Buchanan E, Coffey P, Dixon P, Drummond C, Flanagan M. Low back pain: Early Management of Persistent Non-specific Low Back Pain. [<http://guidance.nice.org.uk/CG88/NICEGuidance/pdf/English>]; 2009.
- [12] Indahl A. Low back pain: diagnosis, treatment, and prognosis. *Scand J Rheumatol* 2004;33:199-209.
- [13] Adams MA. Biomechanics of back pain. *Acupunct Med* 2004;22:178-88.
- [14] Leboeuf-Yde C, Lauritsen JM, Lauritzen T. Why has the search for causes of low back pain largely been nonconclusive? *Spine* 1997;22:877-81.
- [15] Leboeuf-Yde C, Manniche C. Low back pain: Time to get off the treadmill. *J Manipulative Physiol Ther* 2001;24:63-6.

- [16] Allan DB, Waddell G. An historical perspective on low back pain and disability. *Acta Orthop Scand Suppl* 1989;234:1-23.
- [17] Ghormley RK. Low back pain with special reference to the articular facets, with presentation of an operative procedure. *JAMA* 1933;101:1773-7.
- [18] Mixter WJ, Barr J.S. Rupture of the intervertebral disc with involvement of the spinal canal. *N Engl J Med* 1934;2A:210-5.
- [19] Waddell G. A New Clinical-Model for the Treatment of Low-Back-Pain. *Spine* 1987;12:632-44.
- [20] Ramond A, Bouton C, Richard I et al. Psychosocial risk factors for chronic low back pain in primary care--a systematic review. *Fam Pract* 2011;28:12-21.
- [21] Waddell G. *The Back Pain Revolution*. Second ed. Edinburgh: Churchill Livingstone; 2004.
- [22] Kjaer P, Wedderkopp N, Korsholm L, Leboeuf-Yde C. Prevalence and tracking of back pain from childhood to adolescence. *BMC Musculoskelet Dis* 2011;12:98.
- [23] Hartvigsen J, Christensen K. Pain in the back and neck are with us until the end: a nationwide interview-based survey of Danish 100-year-olds. *Spine* 2008;33:909-13.
- [24] Leboeuf-Yde C, Nielsen J, Kyvik KO, Fejer R, Hartvigsen J. Pain in the lumbar, thoracic or cervical regions: do age and gender matter? A population-based study of 34,902 Danish twins 20-71 years of age. *BMC Musculoskel Dis* 2009;10.
- [25] Lidgren L, Smolen J, Bentley G, Delmas P, Woolf A, Åkesson K, Compston J, Thorngren K-G, van Rial P. European Bone and Joint Health Strategies Project. European Action towards Better Musculoskeletal Health. A Public Health Strategy to Reduce the Burden of Musculoskeletal Conditions. Department of Orthopedics, Lund University Hospital; 2004.
- [26] Kjølner M, Juel K, Kamper-Jørgensen F. Folkesundhedsrapporten. <http://www.si-folkesundhed.dk/Udgivelser/B%C3%B8ger%20og%20rapporter/2008/2897%20Folkesundhedsrapporten%202007.aspx>; Statens Institut for Folkesundhed; 2007.
- [27] Leboeuf-Yde C, Fejer R, Nielsen J, Kyvik KO, Hartvigsen J. Consequences of spinal pain: Do age and gender matter? A Danish cross-sectional population-based. *BMC Musculoskel Dis* 2011;12.
- [28] Carey TS, Garrett JM, Jackman AM. Beyond the good prognosis. Examination of an inception cohort of patients with chronic low back pain. *Spine* 2000;25:115-20.
- [29] Coste J, Delecoeuillerie G, Cohen de LA, Le Parc JM, Paolaggi JB. Clinical course and prognostic factors in acute low back pain: an inception cohort study in primary care practice. *BMJ* 1994;308:577-80.
- [30] Schultz IZ, Crook J, Berkowitz J, Milner R, Meloche GR. Predicting return to work after low back injury using the psychosocial risk for occupational disability instrument: A validation study. *J Occup Rehabil* 2005;15:365-76.

- [31] Katz JN. Lumbar disc disorders and low-back pain: Socioeconomic factors and consequences. *J Bone Joint Surg Am* 2006;88A:21-4.
- [32] Frymoyer JW, Catsbaril WL. An Overview of the Incidences and Costs of Low-Back-Pain. *Orthop Clin N Am* 1991;22:263-71.
- [33] Hill JC, Whitehurst DGT, Lewis M et al. Comparison of stratified primary care management for low back pain with current best practice (STarT Back): a randomised controlled trial. *Lancet* 2011;378:1560-71.
- [34] Bjerrum K.M., Davidsen M., Juel K. De samfundsmæssige omkostninger ved rygsygdomme og rygsmerter i Danmark. The National Institute of Public Health, University of Southern Denmark; 2011.
- [35] Hayden JA, Dunn KM, van der Windt DA, Shaw WS. What is the prognosis of back pain? *Best Pract Res Clin Rheumatol* 2010;24:167-79.
- [36] Chou R, Shekelle P. Will this patient develop persistent disabling low back pain? *JAMA* 2010;303:1295-302.
- [37] Steenstra I, Verbeek JH, Heymans MW, Bongers P.M. Prognostic factors for duration of sick leave in patients sick listed with acute low back pain: a systematic review of the literature. *Occup Environ Med* 2005;62:851-60.
- [38] Bakker EW, Verhagen AP, van TE, Lucas C, Koes BW. Spinal mechanical load as a risk factor for low back pain: a systematic review of prospective cohort studies. *Spine* 2009;34:E281-E293.
- [39] Hoogendoorn WE, van Poppel MN, Bongers PM, Koes BW, Bouter LM. Physical load during work and leisure time as risk factors for back pain. *Scand J Work Environ Health* 1999;25:387-403.
- [40] Wai EK, Roffey DM, Bishop P, Kwon BK, Dagenais S. Causal assessment of occupational bending or twisting and low back pain: results of a systematic review. *Spine J* 2010;10:76-88.
- [41] Roffey DM, Wai EK, Bishop P, Kwon BK, Dagenais S. Causal assessment of awkward occupational postures and low back pain: results of a systematic review. *Spine J* 2010;10:89-99.
- [42] Roffey DM, Wai EK, Bishop P, Kwon BK, Dagenais S. Causal assessment of occupational standing or walking and low back pain: results of a systematic review. *Spine J* 2010;10:262-72.
- [43] Roffey DM, Wai EK, Bishop P, Kwon BK, Dagenais S. Causal assessment of occupational sitting and low back pain: results of a systematic review. *Spine J* 2010;10:252-61.
- [44] van Oostrom SH, Verschuren M, de Vet HC, Boshuizen HC, Picavet HS. Longitudinal associations between physical load and chronic low back pain in the general population: the Doetinchem Cohort Study. *Spine* 2012;37:788-96.
- [45] Hartvigsen J, Lings S, Leboeuf-Yde C, Bakketeig L. Psychosocial factors at work in relation to low back pain and consequences of low back pain; a systematic, critical review of prospective cohort studies. *Occup Environ Med* 2004;61:e2.

- [46] Koes BW, van Tulder MW, Ostelo R, Kim BA, Waddell G. Clinical guidelines for the management of low back pain in primary care: an international comparison. *Spine* 2001;26:2504-13.
- [47] Penney JN. A comparison of Australian and European evidence-based guidelines for intervention in acute, non-specific low back pain. *Int J Osteopath Med* 2009;12:63-8.
- [48] Van Tulder M, Becker A, Bekkering T et al. Chapter 3 - European guidelines for the management of acute nonspecific low back pain in primary care. *Eur Spine J* 2006;15:S169-S191.
- [49] Airaksinen O, Brox JJ, Cedraschi C et al. Chapter 4. European guidelines for the management of chronic nonspecific low back pain. *Eur Spine J* 2006;15 Suppl 2:S192-S300.
- [50] Fourney DR, Andersson G, Arnold PM et al. Chronic low back pain: a heterogeneous condition with challenges for an evidence-based approach. *Spine* 2011;36:S1-S9.
- [51] Weiner BK. Spine update: the biopsychosocial model and spine care. *Spine* 2008;33:219-23.
- [52] Norvell DC, Dettori JR, Fehlings MG, Fourney DR, Chapman JR. Methodology for the systematic reviews on an evidence-based approach for the management of chronic low back pain. *Spine* 2011;36:S10-S18.
- [53] Fairbank J, Gwilym SE, France JC et al. The role of classification of chronic low back pain. *Spine* 2011;36:S19-S42.
- [54] Bishop JB, Szpalski M, Ananthraman SK, McIntyre DR, Pope MH. Classification of low back pain from dynamic motion characteristics using an artificial neural network. *Spine* 1997;22:2991-8.
- [55] Dankaerts W, O'Sullivan P, Burnett A, Straker L, Davey P, Gupta R. Discriminating healthy controls and two clinical subgroups of nonspecific chronic low back pain patients using trunk muscle activation and lumbosacral kinematics of postures and movements: a statistical classification model. *Spine* 2009;34:1610-8.
- [56] Dankaerts W, O'Sullivan P. The validity of O'Sullivan's classification system (CS) for a sub-group of NS-CLBP with motor control impairment (MCI): overview of a series of studies and review of the literature. *Man Ther* 2011;16:9-14.
- [57] Magnusson ML, Bishop JB, Hasselquist L, Spratt KF, Szpalski M, Pope MH. Range of motion and motion patterns in patients with low back pain before and after rehabilitation. *Spine* 1998;23:2631-9.
- [58] Marras WS, Parnianpour M, Ferguson SA et al. The Classification of Anatomic-Based and Symptom-Based Low-Back Disorders Using Motion Measure Models. *Spine* 1995;20:2531-46.
- [59] O'Sullivan P. Diagnosis and classification of chronic low back pain disorders: maladaptive movement and motor control impairments as underlying mechanism. *Man Ther* 2005;10:242-55.
- [60] Marras WS, Ferguson SA, Gupta P et al. The quantification of low back disorder using motion measures. Methodology and validation. *Spine* 1999;24:2091-100.
- [61] Evans DW, Lucas N. What is 'manipulation'? A reappraisal. *Man Ther* 2010;15:286-91.

- [62] Farrell JP, Jensen GM. Manual therapy: a critical assessment of role in the profession of physical therapy. *Phys Ther* 1992;72:843-52.
- [63] Harvey E, Burton AK, Moffett JK, Breen A. Spinal manipulation for low-back pain: a treatment package agreed to by the UK chiropractic, osteopathy and physiotherapy professional associations. *Man Ther* 2003;8:46-51.
- [64] van de Veen EA, de Vet HC, Pool JJ, Schuller W, de ZA, Bouter LM. Variance in manual treatment of nonspecific low back pain between orthomanual physicians, manual therapists, and chiropractors. *J Manipulative Physiol Ther* 2005;28:108-16.
- [65] Shekelle PG. Spine Update - Spinal Manipulation. *Spine* 1994;19:858-61.
- [66] Evans DW. Mechanisms and effects of spinal high-velocity, low-amplitude thrust manipulation: Previous theories. *J Manipulative Physiol Ther* 2002;25:251-62.
- [67] Henderson CN. The basis for spinal manipulation: Chiropractic perspective of indications and theory. *J Electromyogr Kinesiol* 2012;22:632-42.
- [68] Pickar JG, Bolton PS. Spinal manipulative therapy and somatosensory activation. *J Electromyogr Kinesiol* 2012.
- [69] Bronfort G, Haas M, Evans R, Kawchuk G, Dagenais S. Evidence-informed management of chronic low back pain with spinal manipulation and mobilization. *Spine J* 2008;8:213-25.
- [70] Evans DW, Breen AC. A biomechanical model for mechanically efficient cavitation production during spinal manipulation: prethrust position and the neutral zone. *J Manipulative Physiol Ther* 2006;29:72-82.
- [71] Herzog W. The biomechanics of spinal manipulation. *J Bodyw Mov Ther* 2010;14:280-6.
- [72] Flynn TW, Fritz JM, Wainner RS, Whitman JM. The audible pop is not necessary for successful spinal high-velocity thrust manipulation in individuals with low back pain. *Arch Phys Med Rehabil* 2003;84:1057-60.
- [73] Flynn TW, Childs JD, Fritz JM. The audible pop from high-velocity thrust manipulation and outcome in individuals with low back pain. *J Manipulative Physiol Ther* 2006;29:40-5.
- [74] Millan M, Leboeuf-Yde C, Budgell B, Amorim MA. The effect of spinal manipulative therapy on experimentally induced pain: a systematic literature review. *Chiropr Man Therap* 2012;20:26.
- [75] Clark BC, Goss DA, Jr., Walkowski S, Hoffman RL, Ross A, Thomas JS. Neurophysiologic effects of spinal manipulation in patients with chronic low back pain. *BMC Musculoskelet Dis* 2011;12:170.
- [76] Bialosky JE, Bishop MD, Robinson ME, Zeppieri G, Jr., George SZ. Spinal manipulative therapy has an immediate effect on thermal pain sensitivity in people with low back pain: a randomized controlled trial. *Phys Ther* 2009;89:1292-303.
- [77] Pickar JG. Neurophysiological effects of spinal manipulation. *Spine J* 2002;2:357-71.

- [78] Bishop MD, Beneciuk JM, George SZ. Immediate reduction in temporal sensory summation after thoracic spinal manipulation. *Spine J* 2011;11:440-6.
- [79] Furlan AD, Imamura M, Dryden T, Irvin E. Massage for low back pain: an updated systematic review within the framework of the Cochrane Back Review Group. *Spine* 2009;34:1669-84.
- [80] Aourell M, Skoog M, Carleson J. Effects of Swedish massage on blood pressure. *Complement Ther Clin Pract* 2005;11:242-6.
- [81] Lindgren L, Rundgren S, Winso O et al. Physiological responses to touch massage in healthy volunteers. *Auton Neurosci* 2010;158:105-10.
- [82] Ernst E. Massage therapy for low back pain: a systematic review. *J Pain Symptom Manage* 1999;17:65-9.
- [83] Chou R, Huffman LH. Nonpharmacologic therapies for acute and chronic low back pain: a review of the evidence for an American Pain Society/American College of Physicians clinical practice guideline. *Ann Intern Med* 2007;147:492-504.
- [84] Bronfort G, Haas M, Evans R, Leininger B, Triano J. Effectiveness of manual therapies: the UK evidence report. *Chiropr Osteopat* 2010;18:3.
- [85] Furlan AD, Yazdi F, Tsertsvadze A et al. A systematic review and meta-analysis of efficacy, cost-effectiveness, and safety of selected complementary and alternative medicine for neck and low-back pain. *Evid Based Complement Alternat Med* 2012;2012:953139.
- [86] Assendelft WJ, Morton SC, Yu EI, Suttrop MJ, Shekelle PG. Spinal manipulative therapy for low back pain. A meta-analysis of effectiveness relative to other therapies. *Ann Intern Med* 2003;138:871-81.
- [87] Rubinstein SM, van Middelkoop M, Assendelft WJJ, de Boer MR, van Tulder MW. Spinal Manipulative Therapy for Chronic Low-Back Pain An Update of a Cochrane Review. *Spine* 2011;36:E825-E846.
- [88] Chou R, Atlas SJ, Stanos SP, Rosenquist RW. Nonsurgical interventional therapies for low back pain: a review of the evidence for an American Pain Society clinical practice guideline. *Spine* 2009;34:1078-93.
- [89] Standaert CJ, Friedly J, Erwin MW et al. Comparative effectiveness of exercise, acupuncture, and spinal manipulation for low back pain. *Spine* 2011;36:S120-S130.
- [90] Slater SL, Ford JJ, Richards MC, Taylor NF, Surkitt LD, Hahne AJ. The effectiveness of sub-group specific manual therapy for low back pain: a systematic review. *Man Ther* 2012;17:201-12.
- [91] van den Borne HW. The patient from receiver of information to informed decision-maker. *Patient Educ Couns* 1998;34:89-102.
- [92] Ivanova JI, Birnbaum HG, Schiller M, Kantor E, Johnstone BM, Swindle RW. Real-world practice patterns, health-care utilization, and costs in patients with low back pain: the long road to guideline-concordant care. *Spine J* 2011;11:622-32.

- [93] Engers A, Jellema P, Wensing M, van der Windt DA, Grol R, van Tulder MW. Individual patient education for low back pain. *Cochrane Database Syst Rev* 2008;CD004057.
- [94] Buchner DM, Bishop J, Brown DR, Fulton JE, Galuska DA. US Department and Health and Human Services. 2008 physical activity guidelines for Americans. [www health gov/PAGuidelines/guidelines/default.aspx](http://www.health.gov/PAGuidelines/guidelines/default.aspx) 2008;
- [95] Barr KP, Griggs M, Cadby T. Lumbar stabilization: core concepts and current literature, Part 1. *Am J Phys Med Rehabil* 2005;84:473-80.
- [96] Hollinghurst S, Sharp D, Ballard K et al. Randomised controlled trial of Alexander technique lessons, exercise, and massage (ATEAM) for chronic and recurrent back pain: economic evaluation. *BMJ* 2008;337:a2656.
- [97] Little P, Lewith G, Webley F et al. Randomised controlled trial of Alexander technique lessons, exercise, and massage (ATEAM) for chronic and recurrent back pain. *BMJ* 2008;337:a884.
- [98] Dunsford A, Kumar S, Clarke S. Integrating evidence into practice: use of McKenzie-based treatment for mechanical low back pain. *J Multidiscip Healthc* 2011;4:393-402.
- [99] Kuukkanen T, Malkia E, Kautiainen H, Pohjolainen T. Effectiveness of a home exercise programme in low back pain: a randomized five-year follow-up study. *Physiother Res Int* 2007;12:213-24.
- [100] Ferreira ML, Ferreira PH, Latimer J et al. Comparison of general exercise, motor control exercise and spinal manipulative therapy for chronic low back pain: A randomized trial. *Pain* 2007;131:31-7.
- [101] Hayden JA, van Tulder MW, Malmivaara AV, Koes BW. Meta-analysis: Exercise therapy for nonspecific low back pain. *Ann Intern Med* 2005;142:765-75.
- [102] Hayden JA, van Tulder MW, Tomlinson G. Systematic review: Strategies for using exercise therapy to improve outcomes in chronic low back pain. *Ann Intern Med* 2005;142:776-85.
- [103] Hayden JA, van Tulder MW, Malmivaara A, Koes BW. Exercise therapy for treatment of non-specific low back pain. *Cochrane Database Syst Rev* 2005;CD000335.
- [104] van Tulder MW, Malmivaara A, Esmail R, Koes BW. Exercise therapy for low back pain. *Cochrane Database Syst Rev* 2000;CD000335.
- [105] Bauman AE. Updating the evidence that physical activity is good for health: an epidemiological review 2000-2003. *J Sci Med Sport* 2004;7:6-19.
- [106] Proper KI, Singh AS, van MW, Chinapaw MJ. Sedentary behaviors and health outcomes among adults: a systematic review of prospective studies. *Am J Prev Med* 2011;40:174-82.
- [107] Griffin DW, Harmon DC, Kennedy NM. Do patients with chronic low back pain have an altered level and/or pattern of physical activity compared to healthy individuals? A systematic review of the literature. *Physiotherapy* 2012;98:13-23.

- [108] Waddell G, Newton M, Henderson I, Somerville D, Main CJ. A Fear-Avoidance Beliefs Questionnaire (FABQ) and the role of fear-avoidance beliefs in chronic low back pain and disability. *Pain* 1993;52:157-68.
- [109] Hallal PC, Victora CG, Azevedo MR, Wells JC. Adolescent physical activity and health: a systematic review. *Sports Med* 2006;36:1019-30.
- [110] Hartvigsen J, Christensen K. Active lifestyle protects against incident low back pain in seniors: a population-based 2-year prospective study of 1387 Danish twins aged 70-100 years. *Spine* 2007;32:76-81.
- [111] Rainville J, Hartigan C, Martinez E, Limke J, Jouve C, Finno M. Exercise as a treatment for chronic low back pain. *Spine J* 2004;4:106-15.
- [112] Merskey H, Bogduk N. Classification of chronic pain. Descriptions of chronic pain syndromes and definitions of pain terms. 2ed edn. ed. Seattle: IASP; 1994.
- [113] Helms JE, Barone CP. Physiology and treatment of pain. *Crit Care Nurse* 2008;28:38-49.
- [114] Meeus M, Nijs J. Central sensitization: a biopsychosocial explanation for chronic widespread pain in patients with fibromyalgia and chronic fatigue syndrome. *Clin Rheumatol* 2007;26:465-73.
- [115] Goffaux P, Redmond WJ, Rainville P, Marchand S. Descending analgesia--when the spine echoes what the brain expects. *Pain* 2007;130:137-43.
- [116] Melzack R, Wall PD. Pain mechanisms: a new theory. *Science* 1965;150:971-9.
- [117] Haldemann S. Principles and practice of chiropractic. Third edition ed. McGraw-Hill Companies; 2005.
- [118] Melzack R, Wall P. Textbook of pain. 4th ed. ed. New York, NY: Churchill livingstone; 1999.
- [119] Staud R. Evidence for shared pain mechanisms in osteoarthritis, low back pain, and fibromyalgia. *Curr Rheumatol Rep* 2011;13:513-20.
- [120] Staud R, Smitherman ML. Peripheral and central sensitization in fibromyalgia: pathogenetic role. *Curr Pain Headache Rep* 2002;6:259-66.
- [121] Bennett RM. Emerging concepts in the neurobiology of chronic pain: evidence of abnormal sensory processing in fibromyalgia. *Mayo Clin Proc* 1999;74:385-98.
- [122] Busch AJ, Barber KA, Overend TJ, Peloso PM, Schachter CL. Exercise for treating fibromyalgia syndrome. *Cochrane Database Syst Rev* 2007;CD003786.
- [123] Cote JN, Hoeger Bement MK. Update on the relation between pain and movement: consequences for clinical practice. *Clin J Pain* 2010;26:754-62.
- [124] Manniche C, Bentzen L, Hesselsoe G, Christensen I, Lundberg E. Clinical-Trial of Intensive Muscle Training for Chronic Low-Back Pain. *Lancet* 1988;2:1473-7.

- [125] Manniche C, Lundberg E, Christensen I, Bentzen L, Hesselsoe G. Intensive dynamic back exercises for chronic low back pain: a clinical trial. *Pain* 1991;47:53-63.
- [126] Mannion AF, Muntener M, Taimela S, Dvorak J. Comparison of three active therapies for chronic low back pain: results of a randomized clinical trial with one-year follow-up. *Rheumatology (Oxford)* 2001;40:772-8.
- [127] Mannion AF, Caporaso F, Pulkovski N, Sprott H. Spine stabilisation exercises in the treatment of chronic low back pain: a good clinical outcome is not associated with improved abdominal muscle function. *Eur Spine J* 2012.
- [128] Standaert CJ, Weinstein SM, Rumpeltes J. Evidence-informed management of chronic low back pain with lumbar stabilization exercises. *Spine J* 2008;8:114-20.
- [129] oktedalen O, Solberg EE, Haugen AH, Opstad PK. The influence of physical and mental training on plasma beta-endorphin level and pain perception after intensive physical exercise. *Stress and Health* 2001;17:121-7.
- [130] Schwarz L, Kindermann W. Changes in beta-endorphin levels in response to aerobic and anaerobic exercise. *Sports Med* 1992;13:25-36.
- [131] Droste C, Greenlee MW, Schreck M, Roskamm H. Experimental pain thresholds and plasma beta-endorphin levels during exercise. *Med Sci Sports Exerc* 1991;23:334-42.
- [132] Rainville J, Jouve CA, Hartigan C, Martinez E, Hipona M. Comparison of short- and long-term outcomes for aggressive spine rehabilitation delivered two versus three times per week. *Spine J* 2002;2:402-7.
- [133] Walsh AJ, Lotz JC. Biological response of the intervertebral disc to dynamic loading. *J Biomech* 2004;37:329-37.
- [134] Wang DL, Jiang SD, Dai LY. Biologic response of the intervertebral disc to static and dynamic compression in vitro. *Spine* 2007;32:2521-8.
- [135] Feine JS, Chapman CE, Lund JP, Duncan GH, Bushnell MC. The perception of painful and nonpainful stimuli during voluntary motor activity in man. *Somatosens Mot Res* 1990;7:113-24.
- [136] Yeziarski RP, Gerhart KD, Schrock BJ, Willis WD. A further examination of effects of cortical stimulation on primate spinothalamic tract cells. *J Neurophysiol* 1983;49:424-41.
- [137] Deyo RA. Diagnostic evaluation of LBP - Reaching a specific diagnosis is often impossible. *Arch Intern Med* 2002;162:1444-7.
- [138] Kjaer P, Leboeuf-Yde C, Korsholm L, Sorensen JS, Bendix T. Magnetic resonance imaging and low back pain in adults: A diagnostic imaging study of 40-year-old men and women. *Spine* 2005;30:1173-80.
- [139] Spitzer WO. Scientific approach to the assessment and management of activity-related spinal disorders: A monograph for clinicians. Report of the Quebec Task Force on Spinal Disorders. *Spine* 1987;12:1-59.

- [140] Atlas SJ, Deyo RA, Patrick DL, Convery K, Keller RB, Singer DE. The Quebec Task Force classification for spinal disorders and the severity, treatment, and outcomes of sciatica and lumbar spinal stenosis. *Spine* 1996;21:2885-92.
- [141] Mooney V. Physical measurements of the lumbar spine. *Phys Med Rehabil Clin N Am* 1998;9:391-410, viii.
- [142] Lehman GJ. Biomechanical assessments of lumbar spinal function. How low back pain sufferers differ from normals. Implications for outcome measures research. Part I: kinematic assessments of lumbar function. *J Manipulative Physiol Ther* 2004;27:57-62.
- [143] McGregor AH, McCarthy ID, Hughes SPF. Motion Characteristics of Normal Subjects and People with Low Back Pain. *Physiotherapy* 1995;81:632-7.
- [144] McGregor AH, Cattermole HR, Hughes SP. Spinal motion in lumbar degenerative disc disease. *J Bone Joint Surg Br* 1998;80:1009-13.
- [145] McGregor AH, Hughes SPF. The potential use of spinal motion as a measure of surgical outcome. *J Back Musculoskelet* 2004;17:77-82.
- [146] McGregor AH, Cattermole HR, Hughes SPF. Global spinal motion in subjects with lumbar spondylolysis and spondylolisthesis - Does the grade or type of slip affect global spinal motion? *Spine* 2001;26:282-6.
- [147] Hazard RG, Bendix A, Fenwick JW. Disability exaggeration as a predictor of functional restoration outcomes for patients with chronic low-back pain. *Spine* 1991;16:1062-7.
- [148] McGregor AH, McCarthy ID, Dore CJ, Hughes SP. Quantitative assessment of the motion of the lumbar spine in the low back pain population and the effect of different spinal pathologies of this motion. *Eur Spine J* 1997;6:308-15.
- [149] Breen AC, Muggleton JM, Mellor FE. An objective spinal motion imaging assessment (OSMIA): reliability, accuracy and exposure data. *BMC Musculoskelet Dis* 2006;7.
- [150] Lund T, Nydegger T, Schlenzka D, Oxland TR. Three-dimensional motion patterns during active bending in patients with chronic low back pain. *Spine* 2002;27:1865-74.
- [151] Kaigle AM, Wessberg P, Hansson TH. Muscular and kinematic behavior of the lumbar spine during flexion-extension. *J Spinal Disord* 1998;11:163-74.
- [152] Ferguson SA, Marras WS, Gupta P. Longitudinal quantitative measures of the natural course of low back pain recovery. *Spine* 2000;25:1950-6.
- [153] Al-Eisa E, Egan D, Deluzio K, Wassersug R. Effects of pelvic skeletal asymmetry on trunk movement - Three-dimensional analysis in healthy individuals versus patients with mechanical low back pain. *Spine* 2006;31:E71-E79.

- [154] Stamos-Papastamos N, Petty NJ, Williams JM. Changes in Bending Stiffness and Lumbar Spine Range of Movement Following Lumbar Mobilization and Manipulation. *J Manipulative Physiol Ther* 2011;34:46-53.
- [155] Dvorak J, Vajda EG, Grob D, Panjabi MM. Normal motion of the lumbar spine as related to age and gender. *Eur Spine J* 1995;4:18-23.
- [156] Gill KP, Callaghan MJ. Intratester and intertester reproducibility of the lumbar motion monitor as a measure of range, velocity and acceleration of the thoracolumbar spine. *Clin Biomech* 1996;11:418-21.
- [157] Lee RY, Laprade J, Fung EH. A real-time gyroscopic system for three-dimensional measurement of lumbar spine motion. *Med Eng Phys* 2003;25:817-24.
- [158] Vanneuville G, Kyndt T, Massaux M et al. Preliminary opto-electronic study on vertebral movement. *Surg Radiol Anat* 1994;16:385-91.
- [159] Moutzouri M, Billis E, Strimpakos N, Kottika P, Oldham JA. The effects of the Mulligan Sustained Natural Apophyseal Glide (SNAG) mobilisation in the lumbar flexion range of asymptomatic subjects as measured by the Zebris CMS20 3-D motion analysis system. *BMC Musculoskel Dis* 2008;9:131.
- [160] Panjabi MM, Oxland TR, Yamamoto I, Crisco JJ. Mechanical-Behavior of the Human Lumbar and Lumbosacral Spine As Shown by 3-Dimensional Load-Displacement Curves. *J Bone Joint Surg Am* 1994;76A:413-24.
- [161] Poitras S, Loisel P, Prince F, Lemaire J. Disability measurement in persons with back pain: A validity study of spinal range of motion and velocity. *Arch Phys Med Rehab* 2000;81:1394-400.
- [162] Gronblad M, Hurri H, Kouri JP. Relationships between spinal mobility, physical performance tests, pain intensity and disability assessments in chronic low back pain patients. *Scand J Rehabil Med* 1997;29:17-24.
- [163] Nattrass CL, Nitschke JE, Disler PB, Chou MJ, Ooi KT. Lumbar spine range of motion as a measure of physical and functional impairment: an investigation of validity. *Clin Rehabil* 1999;13:211-8.
- [164] Sullivan MS, Shoaf LD, Riddle DL. The relationship of lumbar flexion to disability in patients with low back pain. *Phys Ther* 2000;80:240-50.
- [165] Zuberbier OA, Kozlowski AJ, Hunt DG et al. Analysis of the convergent and discriminant validity of published lumbar flexion, extension, and lateral flexion scores. *Spine* 2001;26:E472-E478.
- [166] De Vet HCW, Terwee CB, Mokkink LB, Knol DL. *Measurement in medicin*. Cambridge; 2011.
- [167] Mokkink LB, Terwee CB, Patrick DL et al. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *J Clin Epidemiol* 2010;63:737-45.
- [168] Portney LG, Watkins M.P. *Foundation of clinical research: applications to practice*. Third ed. Upper Saddle River, New Jersey: Pearson Education, Inc.; 2009.

- [169] Aaronson N, Alonso J, Burnam A et al. Assessing health status and quality-of-life instruments: Attributes and review criteria. *Qual Life Res* 2002;11:193-205.
- [170] Bland JM, Altman DG. Statistical-Methods for Assessing Agreement Between 2 Methods of Clinical Measurement. *Lancet* 1986;1:307-10.
- [171] Bland JM, Altman DG. Measuring agreement in method comparison studies. *Stat Methods in Med Res* 1999;8:135-60.
- [172] Christensen HW. Precision and accuracy of an electrogoniometer. *J Manipulative Physiol Ther* 1999;22:10-4.
- [173] Mannion A, Troke M. A comparison of two motion analysis devices used in the measurement of lumbar spinal mobility. *Clin Biomech* 1999;14:612-9.
- [174] Mieritz RM, Bronfort G, Kawchuk G, Breen A., Hartvigsen J. Reliability and measurement error of 3D regional lumbar motion measures: a systematic review. *J Manipulative Physiol Ther* 2012;35:645-56.
- [175] Hayden JA, Cote P, Bombardier C. Evaluation of the quality of prognosis studies in systematic reviews. *Ann Intern Med* 2006;144:427-37.
- [176] Hayden JA, Cote P, Steenstra IA, Bombardier C. Identifying phases of investigation helps planning, appraising, and applying the results of explanatory prognosis studies. *J Clin Epidemiol* 2008;61:552-60.
- [177] Kottner J, Audige L, Brorson S. et al. Guidelines for Reporting Reliability and Agreement Studies (GRRAS) were proposed. *J Clin Epidemiol* 2011;64:96-106.
- [178] van Trijffel E, Anderegg Q, Bossuyt PMM, Lucas C. Inter-examiner reliability of passive assessment of intervertebral motion in the cervical and lumbar spine: A systematic review. *Man Ther* 2005;10:256-69.
- [179] Essendrop M, Maul I, Laubli T, Riihimaki H, Schibye B. Measures of low back function: A review of reproducibility studies. *Clin Biomech* 2002;17:235-49.
- [180] Bossuyt PM, Reitsma JB. The STARD initiative. *Lancet* 2003;361:71.
- [181] Bossuyt PM, Reitsma JB, Bruns DE et al. Towards complete and accurate reporting of studies of diagnostic accuracy: The STARD initiative. *Ann Intern Med* 2003;138:40-4.
- [182] Hollingworth W, Medina LS, Lenkinski RE et al. Interrater reliability in assessing quality of diagnostic accuracy studies using the QUADAS tool: A preliminary assessment. *Academic Radiology* 2006;13:803-10.
- [183] Whiting P, Rutjes AW, Reitsma JB, Bossuyt PM, Kleijnen J. The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. *BMC Med Res Methodol* 2003;3:25.
- [184] Whiting PF, Weswood ME, Rutjes AW, Reitsma JB, Bossuyt PN, Kleijnen J. Evaluation of QUADAS, a tool for the quality assessment of diagnostic accuracy studies. *BMC Med Res Methodol* 2006;6:9.

- [185] Bronfort G, Goldsmith CH, Nelson CF, Boline PD, Anderson AV. Trunk exercise combined with spinal manipulative or NSAID therapy for chronic low back pain: a randomized, observer-blinded clinical trial. *J Manipulative Physiol Ther* 1996;19:570-82.
- [186] Jensen MP, Karoly P, Braver S. The measurement of clinical pain intensity: a comparison of six methods. *Pain* 1986;27:117-26.
- [187] Patrick DL, Deyo RA, Atlas SJ, Singer DE, Chapin A, Keller RB. Assessing Health-Related Quality-Of-Life in Patients with Sciatica. *Spine* 1995;20:1899-908.
- [188] Dopf CA, Mandel SS, Geiger DF, Mayer PJ. Analysis of spine motion variability using a computerized goniometer compared to physical examination. A prospective clinical study. *Spine* 1994;19:586-95.
- [189] Petersen CM, Johnson RD, Schuit D, Hayes KW. Intraobserver and interobserver reliability of asymptomatic subjects' thoracolumbar range of motion using the OSI CA 6000 Spine Motion Analyzer. *J Orthop Sports Phys Ther* 1994;20:207-12.
- [190] Dvorak J, Antinnes JA, Panjabi M, Loustalot D, Bonomo M. Age and gender related normal motion of the cervical spine. *Spine* 1992;17:S393-S398.
- [191] Jordan K, Dziedzic K, Mullis R, Dawes PT, Jones PW. The development of three-dimensional range of motion measurement systems for clinical practice. *Rheumatology* 2001;40:1081-4.
- [192] McGregor AH, McCarthy ID, Hughes SP. Motion characteristics of the lumbar spine in the normal population. *Spine* 1995;20:2421-8.
- [193] Schuit D, Petersen C, Johnson R, Levine P, Knecht H, Goldberg D. Validity and reliability of measures obtained from the OSI CA-6000 Spine Motion Analyzer for lumbar spinal motion. *Man Ther* 1997;2:206-15.
- [194] Feipel V, Rondelet B, Le Pallec JP, Rooze M. Normal global motion of the cervical spine: an electrogoniometric study. *Clin Biomech* 1999;14:462-70.
- [195] Shrout PE, Fleiss JL. Intraclass Correlations - Uses in Assessing Rater Reliability. *Psychol Bull* 1979;86:420-8.
- [196] WHO Expert Committee. Physical status: The use and interpretation of anthropometry. Geneva: WHO; 1995.
- [197] Dworkin RH, Turk DC, Wyrwich KW et al. Interpreting the clinical importance of treatment outcomes in chronic pain clinical trials: IMMPACT recommendations. *J Pain* 2008;9:105-21.
- [198] Ostelo RWJG, Deyo RA, Stratford P et al. Interpreting change scores for pain and functional status in low back pain - Towards international consensus regarding minimal important change. *Spine* 2008;33:90-4.
- [199] Barrett CJ, Singer KP, Day R. Assessment of combined movements of the lumbar spine in asymptomatic and low back pain subjects using a three-dimensional electromagnetic tracking system. *Man Ther* 1999;4:94-9.

- [200] McGregor AH, Hughes SPF. The effect of test speed on the motion characteristics of the lumbar spine during an A-P flexion-extension test. *J Back Musculoskelet* 2000;14:99-104.
- [201] Troke M, Moore AP, Cheek E. Intra-operator and inter-operator reliability of the OSI CA 6000 Spine Motion Analyzer with a new skin fixation system. *Man Ther* 1996;1:92-8.
- [202] Petty NJ. The effect of posteroanterior mobilisation on sagittal mobility of the lumbar spine. *Man Ther* 1995;1:25-9.
- [203] Troke M, Schuit D, Petersen CM. Reliability of lumbar spinal palpation, range of motion, and determination of position. *BMC Musculoskel Dis* 2007;8:103.
- [204] De Vet HCW, Terwee CB, Knol DL, Bouter LM. When to use agreement versus reliability measures. *J Clin Epidemiol* 2006;59:1033-9.
- [205] Shirley FR, O'Connor P, Robinson ME, Macmillan M. Comparison of lumbar range of motion using three measurement devices in patients with chronic low back pain. *Spine* 1994;19:779-83.
- [206] Nevill AM, Atkinson G. Assessing agreement between measurements recorded on a ratio scale in sports medicine and sports science. *Br J Sports Med* 1997;31:314-8.
- [207] Marras WS, Davis KG, Heaney CA, Maronitis AB, Allread WG. The influence of psychosocial stress, gender, and personality on mechanical loading of the lumbar spine. *Spine* 2000;25:3045-54.
- [208] Konstantinou K, Foster N, Rushton A, Baxter D, Wright C, Breen A. Flexion mobilizations with movement techniques: The immediate effects on range of movement and pain in subjects with low back pain. *J Manipulative Physiol Ther* 2007;30:178-85.
- [209] Goodsell M, Lee M, Latimer J. Short-term effects of lumbar posteroanterior mobilization in individuals with low-back pain. *J Manipulative Physiol Ther* 2000;23:332-42.
- [210] Burton AK, Tillotson KM, Edwards VA, Sykes DA. Lumbar sagittal mobility and low back symptoms in patients treated with manipulation. *J Spinal Disord* 1990;3:262-8.
- [211] Lehman GJ, McGill SM. Spinal manipulation causes variable spine kinematic and trunk muscle electromyographic responses. *Clin Biomech* 2001;16:293-9.
- [212] Millan M, Leboeuf-Yde C, Budgell B, Descarreaux M, Amorim MA. The effect of spinal manipulative therapy on spinal range of motion: a systematic literature review. *Chiropr Man Therap* 2012;20:23.
- [213] Taimela S, Harkapaa K. Strength, mobility, their changes, and pain reduction in active functional restoration for chronic low back disorders. *J Spinal Disord* 1996;9:306-12.
- [214] Cook CE. *Othopedic manual therapy: an evidence-based approach*. Upper Saddle River, New Jersey: Pearson Education, Inc.; 2007.
- [215] Ylinen J. *Stretching therapy for sport and manual therapy*. First edition ed. Edinbrugh: Churchill Livingstone Elsevier; 2008.

- [216] Cox ME, Asselin S, Gracovetsky SA et al. Relationship between functional evaluation measures and self-assessment in nonacute low back pain. *Spine* 2000;25:1817-26.
- [217] Kang SW, Lee WN, Moon JH, Chun SI. Correlation of spinal mobility with the severity of chronic lower back pain. *Yonsei Med J* 1995;36:37-44.
- [218] McGregor AH, Dore CJ, McCarthy ID, Hughes SP. Are subjective clinical findings and objective clinical tests related to the motion characteristics of low back pain subjects? *J Orthop Sports Phys Ther* 1998;28:370-7.
- [219] Parks KA, Crichton KS, Goldford RJ, McGill SM. A comparison of lumbar range of motion and functional ability scores in patients with low back pain - Assessment for range of motion validity. *Spine* 2003;28:380-4.
- [220] Cherniack M, Dillon C, Erdil M et al. Clinical and psychological correlates of lumbar motion abnormalities in low back disorders. *Spine J* 2001;1:290-8.
- [221] Dickey JP, Pierrynowski MR, Bednar DA, Yang SX. Relationship between pain and vertebral motion in chronic low-back pain subjects. *Clin Biomech* 2002;17:345-52.
- [222] Leboeuf-Yde C, Larsen K, Ahlstrand I, Volinn E. Coping and back problems: analysis of multiple data sources on an entire cross-sectional cohort of Swedish military recruits. *BMC Musculoskelet Dis* 2006;7:39.
- [223] WHO. Obesity: preventing and managing the global epidemic. Report of a WHO consultation on obesity, Geneva, 3-5 june 1997. Geneva; 1998.
- [224] Ensink FB, Saur PM, Frese K, Seeger D, Hildebrandt J. Lumbar range of motion: influence of time of day and individual factors on measurements. *Spine* 1996;21:1339-43.
- [225] Troke M, Moore AP. The development of a new form of instrument fixation for the OSI CA 6000 spine motion analyzer. *Man Ther* 1995;1:43-6.
- [226] Sjolander P, Michaelson P, Jaric S, Djupsjobacka M. Sensorimotor disturbances in chronic neck pain - Range of motion, peak velocity, smoothness of movement, and repositioning acuity. *Man Ther* 2008;13:122-31.
- [227] Mencken HL. The Divine Afflatus. *New York Evening Mail*; 1917.
- [228] Carey TS, Mielenz TJ. Measuring outcomes in back care. *Spine* 2007;32:S9-14.
- [229] Dworkin R. Objectivity and truth: You'd better believe it. *Philosophy & Public Affairs* 1996;25:87-139.

Manuscript I

LITERATURE REVIEW

RELIABILITY AND MEASUREMENT ERROR OF 3-DIMENSIONAL REGIONAL LUMBAR MOTION MEASURES: A SYSTEMATIC REVIEW

Rune M. Mieritz, DC, MSc,^a Gert Bronfort, DC, PhD,^b Greg Kawchuk, DC, PhD,^c Alan Breen, DC, PhD,^d and Jan Hartvigsen, DC, PhD^e

ABSTRACT

Objective: The purpose of this study was to systematically review the literature on reproducibility (reliability and/or measurement error) of 3-dimensional (3D) regional lumbar motion measurement systems.

Methods: Electronic searches were performed in PubMed, Cumulative Index of the Nursing and Allied Health Literature, Embase, and Mantis databases. To be included, original studies had to report on the reproducibility of a 3D computerized regional lumbar spinal motion analysis system in human subjects. A detailed checklist was developed based on guidelines for reporting reliability and agreement studies, the standards for reporting of diagnostic accuracy, and quality assessment of diagnostic accuracy studies and used for data extraction and quality assessment. The checklist consisted of descriptive items divided into 4 domains: study population, testing circumstances, equipment, and data analysis and presentation. The descriptive items were used as foundation for the quality assessment reflecting the reporting level of the included articles.

Results: A total of 15 articles were included in this study. We found incomplete reporting in 1 or more domains in all articles. A varying amount of measurement error was reported in 8 of the 15 articles. Because of incomplete reporting, these reliability and measurement error estimates are difficult to interpret.

Conclusions: The current literature on the reliability and measurement error of measures created by regional 3D spinal instruments contains uncertainties especially in relevant clinical populations. There is uncertainty with respect to the degree that repeated measurements by 3D regional spinal motion instruments are reproducible. However, limited to the studies where reliability estimates were provided, most instruments used under standardized conditions may be considered reliable enough to be used for research purposes on the group level, but it is uncertain if they can be used on the individual patient level. (*J Manipulative Physiol Ther* 2012;35:645-656)

Key Indexing Terms: *Lumbosacral Region; Back; Range of Motion; Motion; Reproducibility of Results; Review*

Low back pain (LBP) is a major public health problem in the Western world with profound consequences for both individuals and societies.¹⁻³ There are many

theories concerning the etiology of LBP, but in spite of considerable scientific effort, the mechanisms remain largely unknown.^{4,5} Research shows that an estimated

^a PhD Candidate, Institute of Sports Science and Clinical Biomechanics, University of Southern Denmark, Odense, Denmark.

^b Professor, Northwestern Health Sciences University, Minneapolis, MN, and Senior Researcher, Nordic Institute of Chiropractic and Clinical Biomechanics, Odense, Denmark.

^c Associate Professor, Canada Research Chair in Spinal Function, University of Alberta, Edmonton, Alberta, Canada.

^d Professor, Institute for Musculoskeletal Research and Clinical Implementation, Anglo-European College of Chiropractic, Bournemouth, England.

^e Professor, Institute of Sports Science and Clinical Biomechanics, University of Southern Denmark, Odense, Denmark and

Senior Researcher, Nordic Institute of Chiropractic and Clinical Biomechanics, Odense, Denmark.

Submit requests for reprints to: Rune M. Mieritz, DC, MSc, Institute of Sports Science and Clinical Biomechanics, University of Southern Denmark, Campusvej 55, MDK-5230 Odense, Denmark (e-mail: rmieritz@health.sdu.dk).

Paper submitted January 11, 2012; in revised form April 26, 2012; accepted May 17, 2012.

0161-4754/\$36.00

Copyright © 2012 by National University of Health Sciences. <http://dx.doi.org/10.1016/j.jmpt.2012.09.011>

Table 1. Overview of number of citations in each database and number of duplicates relative to PubMed

	PubMed	EM-Base	CINAHL	Mantis	Total	Total no. of unique articles/hits
No. of citations	791	563	308	380	2042	1144
No. of duplicates between database and PubMed		374	238	233		

75% to 85% of the population will have LBP at some time in their lives,⁶ but anatomical diagnosis for LBP conditions is only possible in a minority of the patients.⁷⁻⁹ However, it is assumed that a large proportion of LBP is caused or significantly influenced by biomechanical factors.⁹⁻¹³ The rationale underlying this assumption is that motion characteristics or movement patterns could be potential mediators in the development of LBP, whereas these same kinematics may be altered as a consequence of pain, treatments, or other factors.⁹⁻¹³ Therefore, a variety of technologies have been developed and applied in the attempt to quantify biomechanical characteristics.¹⁴⁻¹⁶ Some technologies have been designed to investigate the motion of each vertebra (segmental motion), that is, digitized videofluoroscopic images¹⁷ or by invasive methods,^{12,18} and others are designed to measure the spine on a regional level, for example, in the entire lumbar spine.^{19,20} Noninvasive, real-time 3-dimensional (3D) regional instruments can provide a quantitative assessment of complex spinal kinematics (eg, coupled motion, combined motion, velocity, acceleration, and smoothness of motion [jerk index]).²¹ Conceptually, these measurement instruments are intended to measure the same construct, however, using different underlying technology, for example, electromagnetic,¹⁶ potentiometric,^{22,23} gyrometric,²⁴ and optoelectronic²⁵ as well as ultrasound pulse²⁶ measuring devices. These noninvasive instruments have no known side effects; compared with roentgenographic analyses, these instruments carry no risk of exposure to radiation. Research suggests that kinematics obtained from noninvasive 3D regional lumbar spinal motion instruments may be of value in generating functional diagnoses, evaluating the mechanisms of therapies, and prescribing specific rehabilitation programs, although no criterion standard exists.²⁰

There are, however, several drawbacks when dealing with instruments attached to the skin surface and measurements in humans. Some of the important factors to consider are measurement error that may originate from the measurement instrument itself; the patients; and the connection between patient and instrument, that is, skin movement artifacts, isolation of pelvic or lumbar segments, or other circumstances under which the measurements take place. A basic premise is that a measurement device should generate reproducible/reliable outcomes when repeated measures are performed on the same subjects under standardized conditions.²⁷ If this is not the case, any use of the instrument in evaluating motion change or generating normative values would be meaningless.²⁸ When reporting

outcomes in reproducibility/reliability studies, it is crucial to provide information detailed enough to understand how the study was conducted and how the results were obtained. Ideally, it should enable others to interpret the results relative to the sample; testing circumstances; design and use of equipment; and, if needed, repeat the study.

Terms such as reliability and reproducibility are routinely used to describe measurements but are often inconsistently defined in the literature. We use the descriptions as presented by De Vet et al²⁹ in which reproducibility is the degree to which repeated measurements provide similar results. Reproducibility can be viewed as an umbrella term for the concepts of agreement and reliability. Agreement focuses on measurement error (“how good is the agreement between repeated measurements”). Reliability refers to the extent in which a measurement instrument can differentiate among individuals or tell individuals apart, despite measurements error.²⁹

The objective of this study was to systematically review the literature and to estimate the quality level of the reporting that addresses the reproducibility (reliability and/or measurement error) of 3D regional lumbar motion measurement systems.

METHODS

Selection of Articles

Relevant articles published before May 2011 were identified by computerized searches. The search strategies were developed in collaboration with a health science research librarian, and electronic searches were performed in PubMed, CINAHL, Embase, and Mantis databases (Table 1). The following search terms were used: (lumbar OR lumbosacral OR lumbal) AND (spinal OR spine OR vertebrae OR vertebral) AND (motion OR biomechanical OR biomechanics OR kinematic OR kinematics) AND (analysis OR analyzing OR analyzer OR measurements OR measuring OR measure OR measures OR assess OR assessed OR assessment) AND (velocity OR ROM OR “range of motion” OR acceleration OR accelerometer) AND (pattern OR patterns OR coupled OR flexion OR extension OR rotation OR rotations OR Lateral). Figure 1 illustrates the screening process.

Inclusion and Exclusion Criteria

To be included, original studies had to report on the reproducibility (test-retest) of a 3D computerized

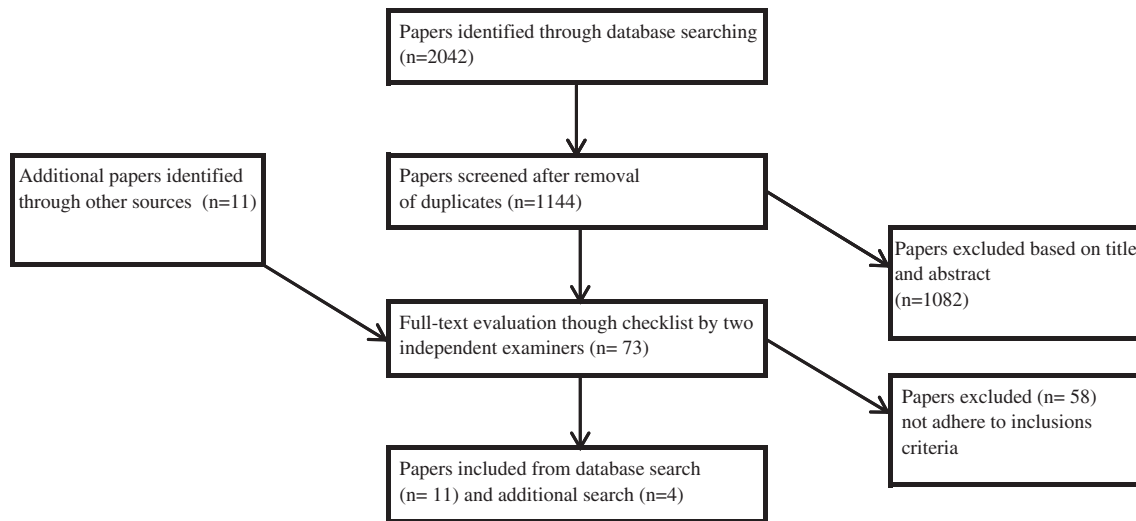


Fig 1. Flow diagram illustrating study selection.

regional lumbar spinal motion analysis system in human subjects. Articles had to be published in peer-reviewed journals as full articles and written in English, Danish, Swedish, or Norwegian. The original studies had to report on the collection of noninvasive 3D data recorded electronically in real-time under standardized conditions. Reproducibility studies using blindfolding or testing of lumbar spine motions while undergoing specific activities such as running and walking were not included in this review. This is because we wanted to evaluate the basis for these types of measurements by evaluating reproducibility protocols of simple and well-defined motion tasks.

Extraction of Information

The checklist was developed based on previous reviews and guidelines for reporting reliability and agreement studies,³⁰⁻³³ the standards for reporting of diagnostic accuracy,^{34,35} and quality assessment of diagnostic accuracy studies.³⁶⁻³⁸ The checklist was designed to guide the systematic extraction of information in a standardized way.^{39,40} This information will be referred to as descriptive items. The descriptive items were then used as foundation for the quality assessment reflecting the reporting level of the included articles. The checklists were divided into 4 domains: study population, testing circumstances, equipment, and data analysis and presentation. A checklist was completed for each included article by 2 of the authors (RMM, JH) independently; the lists were then compared, and disagreements, resolved by discussion. In most cases, descriptive parameters could be checked off as yes or no, and in a few instances, a brief note or numbers were noted as shown in Tables 2, 3, and 4.

The descriptive items under each domain were:

Study Population.

- Anthropometric data: Height and weight.
- Demographic data: Number of participants, age, sex, occupational status, and whether they had LBP. On symptomatic subjects, in addition, measurements of pain, disability, or other reported measures.
- Test population recruitment and inclusion/exclusion criteria.

Testing Circumstances.

- Description of examiners: Professional background and experience level.
- Testing procedures: Time of the day, warm-up, subject instruction, and time interval between trials, data collection location (laboratory, clinic, hospital, etc).
- Attachment level and procedure: Attachment level which described the level of the spine that the instrument was attached to and procedure description related to the protocol used to locate the intended spinal level.
- Measurements performed: Motion being tested in the possible categories flexion (F), extension (E), lateral bending (LB), rotation (R), and other motions (X).
- Testing time duration, time interval between test and retest and information on time consumption.
- Blinding of examiners to clinical information from participants, and their own and others' results.

Equipment.

- Equipment description, the reporting of technological features of the instrument.
- Fixation method including information on which kind of material and technology was used to attach the instrument to the body.
- Instrument precision/accuracy based on its measurement error in relation to a known stable nonhuman quantity.

Table 2. Descriptive matrix of extracted information from each article dealing with an intratester or intertester design and the CA-6000 Spinal Motion Analyser in its original version

First author, year, country, design	Study population		Testing circumstances					Equipment		Data analysis and presentation			
	Study sample (occupation), recruitment, LBP or NLBP ^a	Sex (♂,♀)/N, age (range mean SD)/N ^b	Examiners professional background ^c	Time of the day (Y/N) ^d	Spinal level measured ^e	Movements measured	Time interval between test and retest reported	Description of equipment (Y/N) ^f	Outcome tested	Statistical method	Description of results (Y/N) ^g	Reliability	Agreement
CA 6000													
Dopf, 1994, USA, intra/inter	30 hospital employees, NLBP, Y	♂15 ♀15, N, N, N	N, Y	N, Y, Y, N, N, N	T12/S2, N	F + E + LF + R + X	N/* 2 wk	Y, N, N, N, N	ROM	Pearson, CV	N, Y, NA	<i>r</i> : 0.54-0.97/* <i>r</i> : 0.54-0.84	
Dvorak, 1995, Switzerland, intra/inter	14, NLBP, N/*20, NLBP, N	N, 21-49 35, N, N/* ?♂ ?♀, 31-58 39, N, N	N, Y	Y, Y, Y, N, N, N	T-L/S, Y	F + E + LF + R	3 c*/* consecutively	Y, Y, N, Y, N	ROM	Pearson, CV	N, Y, NA	/* <i>r</i> : 0.72- CV: 5.3-6.7/* 0.851	
McGregor, 1995 UK, intra/inter	10, NLBP, N	9♂ 1♀, 22-54 35, N, N	Physiotherapist and a clinical engineer, Y, N	Y, N, N, N, N, N	T-L/S, N	F + E + LF + R	2-3 d	Y, N, Y, Y, 10 Hz	ROM, VEL	I C C , LOA	Y, Y, N	<i>R</i> : (ROM) 0.39-0.93 <i>R</i> : (VEL) 0.67-0.86/* <i>R</i> : (ROM) 0.33-0.94 <i>R</i> : (VEL) 0.74-0.93 <i>R</i> : 0.89 or higher/* <i>R</i> : 0.85 or higher	LOA (ROM): ≤2.4°± 5.9° LOA (VEL): ≤3.8°/s ±15.4°/s (both intra or inter) SEM (95% CI): ≤3.3 (6.5)/*
Petersen, 1994, USA, intra/inter	21, NLBP, N	8♂ 13♀, 29.7 (5, 6), Y, N	Physiotherapist, Y, N	N, N, N, N, Y, N	T7/S2, Y	F + E + LF + R	2 min/*2 min	Y, Y, Y, Y, 4 Hz	ROM	ICC(2, 1), SEM	Y, Y, Y		CV: 3.3%-7.0% RMSE (SEM): 7.43°-8.6° SEM: 2.30°-3.73°
Petty, 1995, UK, intra	11, NLBP, N	11♀, N, N, N	Therapist, N, N	N, Y, N, N, N, N	T12/S, Y	F + E	3 c*	N, Y, N, Y, N	ROM	CV, RMSE	N, Y, N		
Schuit 1997, USA, intra	30, physical therapy center, LBP*, Y	9♂ 21♀, 23-72 36 (14), Y, Y	N, N	N, Y, Y, N, N, N	T12/S2, Y	F + E + LF + R	2 min	Y, Y, N, Y, 10 Hz	ROM	ICC(2, 1), SEM	N, Y, Y	<i>R</i> : range: 0.875-0.966 mean: 0.9295	

/* separating intratester and intertester design data (intratester data on the left side) or indicating data from 2 different samples. LBP* T12-S1 area symptoms with or without radiating pain into either or both lower extremities, absent neurologic signs, and no previous history of spinal surgery.

F, flexion; E, extension; LF, lateral flexion; R, rotation; X, other motion; N, no/not reported; Y, yes/reported; NA, not applicable; L, lumbar; T, thoracic; S, sacrum; T-L, thoracolumbar; *c, consecutive day; VEL, velocity; LOA, Bland and Altman limits of agreement; RMSE, root mean square error; CV, coefficient of variation; NLBP, non-LBP.

^a Inclusion/exclusion criteria described (Y/N).

^b Height (Y/N), weight (Y/N).

^c Experience/training (Y/N), blinding (Y/N).

^d Warm-up (Y/N), subject instruction (Y/N), testing time duration (Y/N), time interval between trials (Y/N), testing facility (Y/N).

^e Attachment procedure description (Y/N).

^f Fixation method (Y/N), instrument precision/accuracy (Y/N), calibration (Y/N), sampling frequency (Hz/N).

^g Relative to test (Y/N), formula provided (Y/N).

Table 3. Descriptive matrix of extracted information from each article with an intratester or intertester design grouped in different instruments

First author, year, country, design	Study population		Testing circumstances				Equipment		Data analysis and presentation				
	Study sample (occupation), recruitment, LBP or NLBP ^a	Sex (♂,♀)/N, age (range mean SD)/N ^b	Examiners professional background ^c	Time of the day (Y/N) ^d	Spinal level measured ^e	Movements measured	Time interval between test and retest reported	Description of equipment (Y/N) ^f	Outcome tested	Statistical method	Description of results (Y/N) ^g	Reliability	Agreement
Modified CA 6000													
Troke, 1996, UK, intra/inter	11, NLBP, N	7♂ 4♀, 21-35 28 (4, 5), Y, Y	Physiotherapist, Y, N	Y, Y, Y, Y*, N, N	T12/S, Y	F + E + LF + R	4 test days/* 3 operators	Y, Y, N, Y, N	ROM	ICC	N,Y,N	R: 0.795-0.932/* R: 0.822-0.923	
Troke, 2007, UK, intra/inter	22, NLBP, Y	13♂ 9♀, 22-38 (27) 4, 6, Y ,Y	Y, N	N, Y, Y, N, N, N	T12/S, Y	F + E + LF + R	3 d/* 3 operators on 3 d	Y, Y, N, Y, N	ROM	ICC (3, 3)/* ICC (2, 3)	N, Y, Y	R: 0.57-0.99/* R: 0.71-0.82	MVE: <2.5°
Lumbar motion monitor													
Gill, 1996, UK, intra/inter	15, NLBP, N/* 10, NLBP, N	5♂ 10♀, 20–46, N, N/* 3♂ 7♀, N, N	N, N	Y, Y, Y, N, Y, N	T7/S2, Y	F + E + LF + R	48 h	Y, Y, N, Y, 60 Hz	ROM , VEL, ACC	ICC	Y, Y, N	R: (ROM) 0.82-0.87 R: (VEL) 0.61-0.87 R: (ACC) 0.46-0.72/* R: (ROM) 0.93-0.96 R: (VEL) 0.93-0.98 R: (ACC) 0.95-0.98	
Zebris CMS20													
Moutzouri, 2008, Greece, intra	20, NLBP, Y	11♂ 9♀, 19-23 21 (1, 1), Y, Y (NB*)	Physiotherapist, Y, Y	Y, Y, Y, N, N, N	T12/S2, N	F	Same day, 1 wk	Y, Y, N, N, 20 Hz	ROM	ICC (1, 1), SEM	N, Y, Y	R = intraday: 0.89 (day 1) 0.9 (day 2), interday: 0.82	SEM: intraday: 3.3° (day 1) 3.0° (day 2), interday: 4.0°
3 Space Fastrak													
Barret, 1999, Australia, intra	31 local advertising, NLBP, Y/* 16 2 clinics, LBP**, Y	14♂ 17♀, 20-34 26, N, N/* 9♂ 7♀,20-33 26, N, N*	N, N	N, Y, Y, N, N, N	L1/S, Y	F + E + LF + X	5 min	Y, N, Y, N, N	ROM	ICC	N, N, N	R = 0.79-0.92 r = 0.79-0.93	

/* separating intratester and intertester design data (intratester data on the left side) or indicating data from 2 different samples. LBP** currently suffering from LBP and/or lower limb pain related to the lumbar spine NB* all data reported divided in sex.

N*, BMI <30; Y*, 10 min, pr operator; MVE, mean variable error.

^a Inclusion/exclusion criteria described (Y/N).

^b Height (Y/N), weight (Y/N).

^c Experience/training (Y/N), blinding (Y/N).

^d Warm-up (Y/N), subject instruction (Y/N), testing time duration (Y/N), time interval between trials (Y/N), testing facility (Y/N).

^e Attachment procedure description (Y/N).

^f Fixation method (Y/N), instrument precision/accuracy (Y/N), calibration (Y/N), sampling frequency (Hz/N).

^g Relative to test (Y/N), formula provided (Y/N).

- Calibration and sampling frequency.
- Instrument features tested; range of motion (ROM) or higher order kinematics such as velocity or acceleration. In addition reporting of whether the testing speed standardized or subject chosen was noted.

Data Analysis and Presentation.

- Statistical methods used.
- Descriptions of results; the mean and SD or mean difference of the primary movement parameter for each individual testing occasion including intra and intertester where applicable.
- Results presented in relationship to its relative test, for example, intraclass correlation coefficient (ICC) and/or standard error of measurement or Bland and Altman limits of agreement in relation to ROM value in flexion.
- Formula specified where applicable.

Assessment of Quality

The quality assessment was designed to summarize each domain in 1 expression of completeness of reporting based on all of the descriptive items. Categories for each domain were “yes,” “partly,” and “no.” Thus, the best quality assessment evaluation for a study would be yes in all domains. Quality scores were assigned by the 2 authors (RMM, JH) independently when checklists had been completed and scores were subsequently compared and any discrepancies resolved by discussion. The criteria for the scores were:

- “Yes” if the information was found to be complete or very close to complete.
- “Partly” if some but not all information was missing and the reporting in general was considered insufficient.
- “No” when there were major deficiencies or no information reported.

For each of the 4 domains, the specific quality questions were:

- The study sample represents a well-defined population, and description of participants is sufficient to replicate the study group.
- Description of testing procedure and circumstances is sufficient for others to replicate the procedures.
- The description of the equipment is sufficient for others to assess the technology.
- Data presentation is reported in sufficient detail for others to assess the results and the statistical methods.

Analysis and Synthesis

All extracted descriptive information was synthesized (Tables 2, 3, and 4). The information provided by the descriptive extraction was used in the judgments of quality

in relation to each domain and entered into the tables as yes, partly, and no. Finally, the quality of individual articles and the overall evidence was summarized in tables and figures for each group of instruments.

RESULTS

Included Articles

The total number of citations identified in the electronic databases was 2,042 (Table 1). A total of 15 articles were retained after screening of titles and abstracts in relation to the inclusion criteria. Articles failed inclusion due to the following reasons: 507 used other technologies such as roentgenographic analysis or invasive procedures such as spinal surgery^{41,42}; 391 did not use a test-retest or intratester/intertester design; 131 involved other activities such as walking; 71 were not written in English, Danish, Swedish, or Norwegian languages; and 40 were animal studies. For the included articles, reproducibility or reliability of spinal lumbar motion was reported as a sole investigative target in 7 articles,^{23-25,43-46} whereas the remaining 8 articles investigated reproducibility or reliability as part of a study dealing with another primary aim.^{22,26,47-52} There were only minor disagreements between the 2 reviewers with respect to selection of the studies or extraction of data and with respect to the quality assessment judgment. All disagreements were easily resolved by discussion.

Descriptive Items

Overall Study Information. Detailed information on descriptive items for all included studies is provided in Tables 2, 3, and 4. The authors of the original studies applied different designs of repeated measurements; intertester designs were used in 7 articles,^{22,23,45-47,49,52} and intratester designs were used in 11 articles.^{22,23,26,43,45-47,49-52} In the remaining 4 articles, the authors used other designs or other undefined repeated measurement methods on the same subjects under stable conditions.^{24,25,44,48} All studies applying intertester design also contained intratester design.

In total, 5 different 3D motion instruments were examined. The OSI CA-6000 Spine Motion Analyzer (Orthopedic Systems, Inc, Hayward, CA)⁴⁵ was used in its original configuration in 7 articles^{22,44,45,47,49-51} and in a modified version in 2 articles.^{46,52} The Lumbar Spinal Monitor (LMM; Chattecx Corp, Hixson, TN)²³ was tested and reported in 2 articles.^{23,48} The remaining 3 instruments were tested only once.^{24-26,43}

Study Population. In general, the level of reporting regarding study populations was incomplete (Tables 2, 3, and 4). Altogether, a total of 132 men and 129 women plus 34 subjects of unknown sex participated in the 15 studies, the average number of participants was 20 (range, 6-31). A considerable difference in the reporting of age

and anthropometric data between the studies was observed. Both age range and mean were reported in 8 articles.^{22,26,43,44,46,49,51,52} The age ranged from 20 to 72 years with a mean of 27.5 ± 8.3 SD, and the body mass index (BMI) ranged from 21 to 27 with a mean of 23.9 ± 1.8 SD, based on information when this was provided (Tables 2, 3, and 4). In 7 articles, no anthropometric information was reported.^{22,23,25,44,47,49,50} Specific inclusion/exclusion criteria were described in only 4 of the 15 articles.^{26,43,47,51,52} In 12 articles, pain-free subjects were included, and in 3, patients had LBP.^{43,44,51}

Testing Circumstances. The general level of reporting regarding the testing protocol varied considerably between articles and was mostly incomplete (Tables 2, 3, and 4). The educational background of the examiners were described in 5 articles of the 15^{26,45,46,49,50} (Tables 2, 3, and 4), and the experience of the examiners in working with elements of the protocol was reported in 5 articles, for example, familiar with skin surface marking techniques for spine and trained in use of equipment to 4 hours training with the equipment.^{26,45,46,49,52} Information on blinding of examiners was provided in 3 of the 15 articles.^{22,26,47} Time interval between test and retest varied from around a month⁴⁸ to 1 or 2 weeks,^{26,47} days,^{22,23,46,49,50,52} and same day^{24,26,43,45,51} and was not reported in 2 articles.^{25,44}

Equipment. In general, description of equipment was complete when provided; a description of instrument accuracy or/and precision was not provided in 9 articles.^{22,23,25,26,46,47,50-52} The instrument outcome reported was ROM in 13 articles,^{22,23,26,43-52} ROM and higher order kinematics in 4,^{23,44,48,49} and ROM and motion patterns in 1.²⁴ In 1 article, no outcome measure was reported.²⁵

Data Analysis and Presentation. Data presentation was sufficient to assess analysis adequacy in 4 of the 15 articles.^{23,44,45,49}

A range of data analysis and statistical methods were applied in the articles addressing reliability (Tables 2, 3, and 4). Intraclass correlation coefficients were reported with formulae specified in 4 articles.^{26,45,51,52} The ICCs were reported without the formulas specified in 4 studies.^{23,43,46,49} Cronbach α was used in 1 study⁴⁸; correlation coefficients were used in 3 articles.^{22,47,53} In 1 article, no statistical methods were reported.²⁵

Different methods were also used in the reporting of agreement parameters. The Bland and Altman mean difference technique (mean difference \pm SD) was reported in one article,⁴⁹ and mean difference \pm 2 SD in another.⁴⁴ Standard error of measurement were reported in 3 articles,^{26,45,51} and the coefficient of variation, in 3.^{22,47,50}

Quality Assessment. Figure 2 contains the summarized quality assessments of all articles relative to each domain. In relation to the domains “study population” and “testing circumstances,” the articles scored “yes” due to complete or very close to complete reporting in only 33% and 20%,

respectively. The domain “equipment” had the highest completeness of reporting with 73% of the articles reporting adequately. Only 13% of the articles were assigned “yes” in relation to reporting of “data presentation.”

Table 5 contains the articles and their individual quality assessments grouped by instrument.

The Reported Level of Reproducibility

In general, no obvious pattern between the quality of reporting and the studies reproducibility results was found.

Reliability. Intraclass correlation coefficient values for ROM parameters was found to be 0.75 or greater in all motion directions in 6 articles using intratester design^{23,26,43,45,46,51} and in 3 studies using intertester design.^{23,45,46} In 2 studies, lower coefficients were found in some of the motion directions.^{49,52}

Two studies examined higher order kinematics.^{23,49} McGregor et al⁴⁹ reported ICC velocity parameters ranging from 0.67 to 0.86 using intratester design and 0.74 to 0.98 using intertester design. Gill and Callaghan²³ found ICC values ranging from 0.61 to 0.87 using intratester design and 0.93 or more using intertester design. In addition Gill and Callaghan reported acceleration parameters ranging from 0.46 to 0.72 using intratester design and 0.95 to 0.97 using intertester design.²³

Agreement. Standard error of measurement values based on ROM data were presented in 3 articles^{26,45,51} ranging from 2.3° to 6.5° and reported as root mean square error (7.43°–8.6°) in 1 article.⁵⁰ Bland and Altman mean difference limits of agreement technique were reported in 2 articles based on ROM and velocity data.^{44,49} Intratester and intertester (mean difference \pm SD) values based on ROM values was ranging from -1.8° to $2.4^\circ \pm 2.4^\circ$ to 5.9° , and velocity values was ranging from 0.1° per second to 3.8° per second $\pm 5.3^\circ$ per second to 15.4° per second.⁴⁹ In the other study, intratester and intertester (mean difference \pm 2 SDs) values based on ROM values was ranging from -1.8° to $2.2^\circ \pm 5^\circ$ to 14.2° , and velocity values ranged from -49.4° to $8.3^\circ \pm 7.8^\circ$ to 131.0° .⁴⁴

Reproducibility results from all studies are presented in Tables 2, 3, and 4.

DISCUSSION

To our knowledge, this is the most comprehensive overview and critical appraisal of the literature addressing the reproducibility of regional lumbar spinal motion measurements to date. The results of this review indicate that many different 3D motion assessment instruments can obtain acceptable reliability coefficients. Unfortunately, these estimates are difficult to interpret due to incomplete reporting with respect to how these studies were performed and a lack of reporting of agreement parameters. Consequently, the exact meaning and usefulness of this

Table 4. Descriptive matrix of extracted information from each article with other or undefined test-retest design grouped in different instruments

First author, year, country	Study population		Testing circumstances					Equipment		Data analysis and presentation			
	Study sample (occupation), recruitment, LBP or NLBP ^a	Sex (♂,♀)/ N, age (range mean SD)/N ^b	Examiners professional background ^c	Time of the day (Y/N) ^d	Spinal level measured ^e	Movements measured	Time interval between test and retest reported	Description of equipment (Y/N) ^f	Outcome tested	Statistical method	Description of results (Y/N) ^g	Reliability	Agreement
CA 6000													
McGregor, 2000, UK	15 college staff, NLBP, N/* 15 hospital recruitment, LBP*, N	12♂, 3♀, 23-45 33, 3 6, 3, N, N/* 6♂, 9♀, 33-70 58 16, 4, N, N	N, N	N, Y, N, N, N, N	TH-L/S, N	F + E	N	Y, Y, Y, N, 10 Hz	ROM, Vel NB*	LOA	Y, Y, NA		ROM F −1.2° ± 4.1° E 0.5° ± 8.0° VEL F −16.9° ± 37.4° E 4.1° ± 21.8°/* ROM: F 0.15° ± 5.0° E −1.6° ± 10.0° VEL F −4.4° ± 9.6° E −4° ± 7.8°
Lumbar motion monitor													
Marras, 1995	20, NLBP, N	10♂, 10♀, 27 4, 8, Y, Y	N, N	N, N, Y, N, N, N	NR, N	F + E + X	Once a week in 5 wk	Y, Y, N, N, 60 Hz	ROM, VEL, ACC NB**	Cronbach α	N, Y, NA	ROM 0.47-0.96 (VEL) 0.57- 0.96 (ACC) 0.61-0.96	
IS 300 Pro, gyroscope													
Lee, 2003	19, NLBP, N	15♂, 4♀, 22 5, Y, N	N, N	N, N, Y, N, Y, N	L1/S, N	F + E + LF + R	Immediately	Y, Y, Y, N 200 Hz	ROM, pattern	CMC	N, N, NA	CMC = 0.972- 0.991	N*
Mac Reflex													
Vanneuville, 1994	6, NLBP, N	4♂, 2♀, 20-35, N, N	N, N	N, N, N, N, N, N	T1, T7, T12, L1, L3, L5, S1, N	F + E + LR + R	N	N, N, N, N, N	N*	N	N, N, N	N*	N*

/* separating intratester and intertester design data (intratester data on the left side) or indicating data from 2 different samples. LBP* (radiologic presentation of central spinal stenosis), NB*, preferred; +, standardized slow or fast speed (only preferred presented in table); NB**, controlled motion speed; C, Cronbach α; CMC, coefficient of multiple correlation; N*, only graphical presentations.

^a Inclusion/exclusion criteria described (Y/N).

^b Height (Y/N), weight (Y/N).

^c Experience/training (Y/N), blinding (Y/N).

^d Warm-up (Y/N), subject instruction (Y/N), testing time duration (Y/N), time interval between trials (Y/N), testing facility (Y/N).

^e Attachment procedure description (Y/N).

^f Fixation method (Y/N), instrument precision/accuracy (Y/N), calibration (Y/N), sampling frequency (Hz/N).

^g Relative to test (Y/N), formula provided (Y/N).

body of literature are not known, and performance of these instruments in clinical practice can be questioned. Furthermore, we did not find a correlation between quality of reporting and the reproducibility results. Suggesting that these studies are quite similar, but due to variation in the level of reporting, this is difficult to know. We found no difference in the level of reproducibility between subjects groups (LBP vs NLBP). However, in general, intratester and short time interval between measurements tended to result in higher reliability coefficients (Tables 2, 3, and 4).

There was considerable variation in the quality and completeness of reporting between the domains. For the domain “study population,” it appears that mostly convenience samples,⁵⁴ that is, colleagues or students of low BMI without LBP, were used; therefore, results presented pertain only to this or similar populations and not to clinical populations. For the domain “testing circumstances,” information needed to be aware of the fundamental controllable procedures of the study was often not reported. The level of reproducibility on different instruments may be influenced by study population characteristics and/or different testing circumstances. With improved reporting in future studies, it may be possible for other researchers to replicate the methods on different samples or devices. This is necessary to enhance our understanding of mechanisms affecting measurement variability and thereby the clinical usefulness of these instruments. The domain “equipment” contained the technical description of the devices used and was the domain with the greatest completion of reporting. However, only a few aspects of advanced lumbar spine kinematics were evaluated in only 5 of the 15 articles.^{23,24,44,48,49} Parameters other than ROM, for example, higher order kinematic motion, are potentially clinically relevant,^{20,21} and therefore, we suggest that future studies examine the reproducibility of kinematic parameters such as velocity, coupled or smoothness of motion in addition to ROM.

The importance of rest-retest or reproducibility studies as a foundation in this field of science is increasingly accepted; however, the standards for statistical analysis and reporting are still evolving.^{30,55} This evolution may explain some of the variation in reporting and choice of statistics in some of the earliest studies included in this review where statistics such as Pearson correlation coefficient and coefficient of variation were used.^{22,47,50} Pearson correlation coefficient is considered to be a reliability measure by some,^{22,47} but it is different from the ICC and not appropriate to use in reliability studies today.^{28,29} Pearson correlation coefficient is based on regression analysis and will usually be higher than the true reliability.²⁸ However, when the predominant source of error is due to random variation, they provide similar values as seen in the study by Barrett et al.⁴³ Coefficient of variation is used by the authors as an outcome in 3 studies.^{22,47,50} This expression

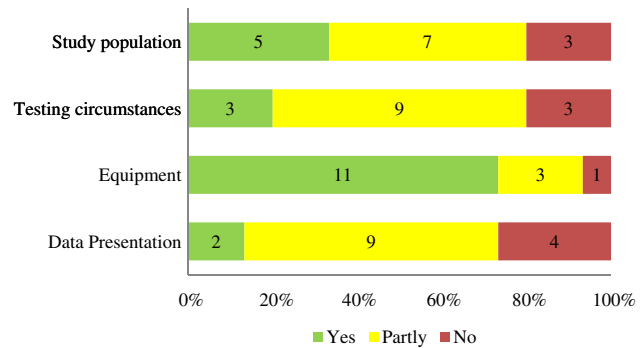


Fig 2. Overview of results of the quality assessment in each domain. (Color version of figure is available online.)

has some disadvantages for this type of data, for example, only random error is assessed and it is independent of the unit of observation, which makes it more difficult to relate to as an agreement parameter in the clinical setting.

The ICC is a unitless index based on the ratio of the within-to-between subject differences and ranges from 0.00 to 1.00, with values closer to 1.00 representing stronger reliability. For the ICC, there are no standard values for acceptable reliability. As a guideline, Portney and Watkins⁵⁶ suggest that values above 0.75 are indicative of good reliability and those below 0.75 poor to moderate. Aaronson et al³³ recommend that minimal standards for reproducibility coefficients be 0.70 for group comparisons and 0.90 to 0.95 for individual measurements over time. Intraclass correlation coefficients were the most frequently applied statistic, and these values are potentially informative but only if the appropriate formula for the particular research question is used. We found that half of the studies did not specify the formula when they used ICC, which makes it problematic to interpret the values. In addition, without underlying variance components, it cannot be known how much influence measurement error has compared with the influence of the variability between the subjects. Measurement error (eg, standard error of measurement) can be derived from the ICC formulae²⁸; however, as pointed out by De Vet et al,²⁹ this is only possible if all components of the ICC formulae is presented by the author and none of the authors of the included studies presented all components of the ICC formulae. Standard error of measurement, 95% limits of agreement, and the smallest detectable change are informative agreement parameters when these instruments are used to detect treatment effects beyond measurement error. The smallest detectable change can be calculated based on standard error of measurement.²⁹ These estimates give a direct measure of how much change or “signal” is needed to exceed the error or “noise” in a specific scientific or clinical setup. These agreement parameters were only very sparsely reported in the literature and when reported any explanation of the

Table 5. Overview of all included studies and its individual quality ratings yes, partly, and no

Quality assessment	Study population	Testing circumstances	Equipment	Data presentation
CA 6000				
Dopf (1994)	Partly	Partly	Partly	Partly
Dvorak (1995)	No	Partly	Yes	Partly
McGregor (1995)	Partly	Partly	Yes	Partly
McGregor (2000)	Partly	No	Yes	Yes
Petersen (1994)	Partly	Partly	Yes	Yes
Petty (1995)	No	Partly	Partly	No
Schuit (1997)	Yes	Partly	Yes	Partly
Modified CA 6000				
Troke (1996)	Yes	Yes	Yes	Partly
Troke (2007)	Yes	Partly	Yes	Partly
3Space Fastrak				
Barret (1999)	Partly	Partly	Partly	No
Lumbar motion monitor				
Gill (1996)	Partly	Yes	Yes	Partly
Marras (1995)	Yes	No	Yes	Partly
Zebis CMS20				
Moutzouri (2008)	Yes	Yes	Yes	Partly
Mac Reflex				
Vanneuville (1994)	No	No	No	No
IS 300 Pro, gyroscope				
Lee (2003)	Partly	Partly	Yes	No

interpretation was absent. We therefore recommend that future articles report at least one of these aspects in detail. It has been described how to do sample size estimation when designing a study to estimate the standard error of measurement.⁵⁷ However, none of the included reliability studies in this systematic review did report the use of sample size calculations. This might be considered when designing future reliability studies.

We found overall ICC based reliability coefficients above 0.7 for most measurements parameters indicating that these instruments used on humans under standardized conditions may be reliable enough to be used for research purposes, for example, measure change in back motion over time on the group level, but it is uncertain if any can be used on the individual patients' level (Tables 2, 3, and 4).

Future investigations should perform test-retest studies on motion parameters other than ROM in different populations, that is, persons with and without LBP and obese and lean persons, to mention the most obvious ones and agreement parameters should be provided and interpreted.^{43,51}

Limitations

A quality assessment tool is a required element in critical reviews to limit bias and improve reviewer consistency.

Unfortunately, no tool for our unique context had been developed. Therefore, a comprehensive checklist was developed based on reviews on the same topic and guidelines.³⁰⁻³⁸ The checklist was designed to guide the systematic extraction of information in a standardized and reproducible way. All descriptive results from the primary studies were presented in detail in Tables 2, 3, and 4 enabling researchers to redo and test the data extraction. The checklist was completed for each included article by 2 of the authors (RMM, JH) independently to enhance the validity of the study. The validity and reliability of this instrument have not been formally tested and may, therefore, potentially be biased.

The number of studies and the methodological and reporting issues is large, and therefore, we did not enter into assessment and discussions about the validity of these measurements. We propose that our checklist is used as a guide for reporting by other researchers when reporting reproducibility studies in this area. The usefulness of these instruments for the evaluation of individual patients in clinical practice is still unclear, and therefore, they must be used with caution. However, to standardize protocols and document procedures in clinics, relevant information is provided in this review.

A number of reporting biases may affect our results. We weighted the perceived chances of finding relevant articles against the time resources available and decided not to include the gray literature, for example, unpublished studies (potential "publication bias") and not to perform a hand-search of selected journals in addition to the primary search. Furthermore, there may be relevant articles published in other languages, which are not included. We did find 11 studies for full-text evaluation based on careful scrutiny of reference lists in included articles, and 4 of these were finally included. Reproducibility studies using blindfolding or testing lumbar spinal motion during specific activities such as running and walking were also not included in this review.

CONCLUSION

According to the literature reviewed in this study, acceptable reliability coefficients can be obtained by different instruments used for the evaluation of 3D lumbar regional spinal motion. Unfortunately, because of incomplete reporting in relation to study samples, testing circumstances, testing protocol, and statistics, these reliability estimates are difficult to interpret. Therefore, there is uncertainty with respect to the degree that repeated measurements by 3D regional spinal motion instruments are reproducible. The results of this review indicate that most instruments used under standardized conditions may be considered reliable enough to be used for research purposes on the group level, but it is uncertain if any can be used on the individual patients' level. However,

Practical Applications

- Acceptable reliability coefficients (ICC >0.7) have been obtained by many different 3D regional lumbar measurement instruments on a variety of motion parameters, but because of incomplete reporting, these reliability estimates are difficult to interpret.
- Three-dimensional regional motion measurements appear to be reliable for measurements at the group but not at the individual patient level.

this conclusion is limited to the studies where reliability estimates was provided.

ACKNOWLEDGMENT

We acknowledge Dr Roni Evans for assistance with manuscript preparation and Dr Johan Wallin for expert research librarian assistance.

FUNDING SOURCES AND POTENTIAL CONFLICTS OF INTEREST

This project was supported by the Danish Agency for Science, Technology, and Innovation, Danish Chiropractors Research Foundation, and the University of Southern Denmark. No conflicts of interest were reported for this study.

REFERENCES

- Andersson GBJ. Epidemiology of low back pain. *Acta Ortho Scand* 1998;69:28-31.
- Manek NJ, MacGregor AJ. Epidemiology of back disorders: prevalence, risk factors, and prognosis. *Curr Opin Rheumatol* 2005;17:134-40.
- Woolf AD, Pfleger B. Burden of major musculoskeletal conditions. *B World Health Organ* 2003;81:646-56.
- Leboeuf-Yde C, Lauritsen JM, Lauritzen T. Why has the search for causes of low back pain largely been nonconclusive? *Spine* 1997;22:877-81.
- Leboeuf-Yde C, Manniche C. Low back pain: Time to get off the treadmill. *J Manipulative Physiol Ther* 2001;24:63-6.
- Andersson GBJ. Epidemiological features of chronic low-back pain. *Lancet* 1999;354:581-5.
- Deyo RA. Diagnostic evaluation of LBP - Reaching a specific diagnosis is often impossible. *Arc Int Med* 2002;162:1444-7.
- Kjaer P, Leboeuf-Yde C, Korsholm L, Sorensen JS, Bendix T. Magnetic resonance imaging and low back pain in adults: A diagnostic imaging study of 40-year-old men and women. *Spine* 2005;30:1173-80.
- O'Sullivan P. Diagnosis and classification of chronic low back pain disorders: maladaptive movement and motor control impairments as underlying mechanism. *Man Ther* 2005;10:242-55.
- Adams MA. Biomechanics of back pain. *Acupunct Med* 2004;22:178-88.
- Bishop JB, Szpalski M, Ananthraman SK, McIntyre DR, Pope MH. Classification of low back pain from dynamic motion characteristics using an artificial neural network. *Spine* 1997;22:2991-8.
- Kaigle AM, Wessberg P, Hansson TH. Muscular and kinematic behavior of the lumbar spine during flexion-extension. *J Spinal Disord* 1998;11:163-74.
- Panjabi MM, Oxland TR, Yamamoto I, Crisco JJ. Mechanical-behavior of the human lumbar and lumbosacral spine as shown by 3-dimensional load-displacement curves. *J Bone Joint Surg Am* 1994;76A:413-24.
- Kuo YL, Tully EA, Galea MP. Video analysis of sagittal spinal posture in healthy young and older adults. *J Manipulative Physiol Ther* 2009;32:210-5.
- De Carvalho DE, Soave D, Ross K, Callaghan JP. Lumbar spine and pelvic posture between standing and sitting: a radiologic investigation including reliability and repeatability of the lumbar lordosis measure. *J Manipulative Physiol Ther* 2010;33:48-55.
- Stamos-Papastamos N, Petty NJ, Williams JM. Changes in bending stiffness and lumbar spine range of movement following lumbar mobilization and manipulation. *J Manipulative Physiol Ther* 2011;34:46-53.
- Breen AC, Muggleton JM, Mellor FE. An objective spinal motion imaging assessment (OSMIA): reliability, accuracy and exposure data. *BMC Musculoskelet Disord* 2006;4:7.
- Lund T, Nydegger T, Schlenzka D, Oxland TR. Three-dimensional motion patterns during active bending in patients with chronic low back pain. *Spine* 2002;27:1865-74.
- Ferguson SA, Marras WS, Gupta P. Longitudinal quantitative measures of the natural course of low back pain recovery. *Spine* 2000;25:1950-6.
- Lehman GJ. Biomechanical assessments of lumbar spinal function. How low back pain sufferers differ from normals. Implications for outcome measures research. Part I: kinematic assessments of lumbar function. *J Manipulative Physiol Ther* 2004;27:57-62.
- Sjolander P, Michaelson P, Jaric S, Djupsjobacka M. Sensorimotor disturbances in chronic neck pain—range of motion, peak velocity, smoothness of movement, and repositioning acuity. *Man Ther* 2008;13:122-31.
- Dvorak J, Vajda EG, Grob D, Panjabi MM. Normal motion of the lumbar spine as related to age and gender. *Eur Spine J* 1995;4:18-23.
- Gill KP, Callaghan MJ. Intratester and intertester reproducibility of the lumbar motion monitor as a measure of range, velocity and acceleration of the thoracolumbar spine. *Clin Biomech* 1996;11:418-21.
- Lee RY, Laprade J, Fung EH. A real-time gyroscopic system for three-dimensional measurement of lumbar spine motion. *Med Eng Phys* 2003;25:817-24.
- Vanneuville G, Kyndt T, Massaux M, Harmand Y, Garcier JM, Monnet JP, et al. Preliminary opto-electronic study on vertebral movement. *Surg Radiol Anat* 1994;16:385-91.
- Moutzouri M, Billis E, Strimpakos N, Kottika P, Oldham JA. The effects of the Mulligan Sustained Natural Apophyseal Glide (SNAG) mobilisation in the lumbar flexion range of asymptomatic subjects as measured by the Zebris CMS20 3-D motion analysis system. *BMC Musculoskelet Disord* 2008;9:131.
- Chinn S. The assessment of methods of measurement. *Stat Med* 1990;9:351-62.
- Streiner DL, Norman GR. *Health Measurement Scales. A practical guide to their development and use.* 4th ed. New York: Oxford University Press Inc.; 2008.
- De Vet HCW, Terwee CB, Knol DL, Bouter LM. When to use agreement versus reliability measures. *J Clin Epidemiol* 2006;59:1033-9.

30. Kottner J, Audige L, Brorson S, Donner A, Gajewski BJ, Hróbjartsson A, et al. Guidelines for Reporting Reliability and Agreement Studies (GRRAS) were proposed. *J Clin Epidemiol* 2011;64:96-106.
31. van Trijffel E, Anderegg Q, Bossuyt PMM, Lucas C. Inter-examiner reliability of passive assessment of intervertebral motion in the cervical and lumbar spine: a systematic review. *Man Ther* 2005;10:256-69.
32. Essendrop M, Maul I, Laubli T, Riihimäki H, Schibye B. Measures of low back function: a review of reproducibility studies. *Clin Biomech* 2002;17:235-49.
33. Aaronson N, Alonso J, Burnam A, Lohr KN, Patrick DL, Perrin E, et al. Assessing health status and quality-of-life instruments: attributes and review criteria. *Qual Life Res* 2002;11:193-205.
34. Bossuyt PM, Reitsma JB. The STARD initiative. *Lancet* 2003;361:71.
35. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig LM, et al. Towards complete and accurate reporting of studies of diagnostic accuracy: the STARD initiative. *Ann Intern Med* 2003;138:40-4.
36. Hollingworth W, Medina LS, Lenkinski RE, Shibata DK, Bernal B, Zurakowski D, et al. Interrater reliability in assessing quality of diagnostic accuracy studies using the QUADAS tool: a preliminary assessment. *Acad Radiol* 2006;13:803-10.
37. Whiting P, Rutjes AW, Reitsma JB, Bossuyt PM, Kleijnen J. The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. *BMC Med Res Methodol* 2003;10:3-25.
38. Whiting PF, Weswood ME, Rutjes AW, Reitsma JB, Bossuyt PN, Kleijnen J. Evaluation of QUADAS, a tool for the quality assessment of diagnostic accuracy studies. *BMC Med Res Methodol* 2006;6:9.
39. Hayden JA, Cote P, Bombardier C. Evaluation of the quality of prognosis studies in systematic reviews. *Ann Intern Med* 2006;144:427-37.
40. Hayden JA, Cote P, Steenstra IA, Bombardier C. Identifying phases of investigation helps planning, appraising, and applying the results of explanatory prognosis studies. *J Clin Epidemiol* 2008;61:552-60.
41. Lim MR, Loder RT, Huang RC, Lyman S, Zhang K, Sama A, et al. Measurement error of lumbar total disc replacement range of motion. *Spine* 2006;31:E291-7.
42. Chen C-S, Cheng C-K, Liu C-L. A biomechanical comparison of posterolateral fusion and posterior fusion in the lumbar spine. *J Spinal Disord Tech* 2002;15:53-63.
43. Barrett CJ, Singer KP, Day R. Assessment of combined movements of the lumbar spine in asymptomatic and low back pain subjects using a three-dimensional electromagnetic tracking system. *Man Ther* 1999;4:94-9.
44. McGregor AH, Hughes SPF. The effect of test speed on the motion characteristics of the lumbar spine during an A-P flexion-extension test. *J Back Musculoskelet* 2000;14:99-104.
45. Petersen CM, Johnson RD, Schuit D, Hayes KW. Intraobserver and interobserver reliability of asymptomatic subjects' thoracolumbar range of motion using the OSI CA 6000 Spine Motion Analyzer. *J Orthop Sports Phys Ther* 1994;20:207-12.
46. Troke M, Moore AP, Cheek E. Intra-operator and inter-operator reliability of the OSI CA 6000 Spine Motion Analyzer with a new skin fixation system. *Man Ther* 1996;1:92-8.
47. Dopf CA, Mandel SS, Geiger DF, Mayer PJ. Analysis of spine motion variability using a computerized goniometer compared to physical examination. A prospective clinical study. *Spine* 1994;19:586-95.
48. Marras WS, Parnianpour M, Ferguson SA, Kim JY, Crowell RR, Bose S, et al. The classification of anatomic-based and symptom-based low-back disorders using motion measure models. *Spine* 1995;20:2531-46.
49. McGregor AH, McCarthy ID, Hughes SP. Motion characteristics of the lumbar spine in the normal population. *Spine* 1995;20:2421-8.
50. Petty NJ. The effect of posteroanterior mobilisation on sagittal mobility of the lumbar spine. *Man Ther* 1995;1:25-9.
51. Schuit D, Petersen C, Johnson R, Levine P, Knecht H, Goldberg D. Validity and reliability of measures obtained from the OSI CA-6000 Spine Motion Analyzer for lumbar spinal motion. *Man Ther* 1997;2:206-15.
52. Troke M, Schuit D, Petersen CM. Reliability of lumbar spinal palpation, range of motion, and determination of position. *BMC Musculoskelet Disord* 2007;8:103.
53. Lee R. Measurement of movements of the lumbar spine. *Physiother Theory Pract* 2002;18:159-64.
54. Lunsford TR, Lunsford BR. The research sample, part I: sampling. *J Prosthet Orthot* 1995;7:105-12.
55. De Vet HCW, Terwee CB, Bouter LM. Current challenges in clinimetrics. *J Clin Epidemiol* 2003;56:1137-41.
56. Portney LG, Watkins MP. Foundation of clinical research: applications to practice. 3rd ed. Upper Saddle River, NJ: Pearson Education, Inc.; 2009.
57. Stratford PW, Goldsmith CH. Use of the standard error as a reliability index of interest: an applied example using elbow flexor strength data. *Phys Ther* 1997;77:745-50.

Manuscript II

Reliability and measurement error of regional lumbar motion parameters in 220 chronic low back pain patients

Rune M. Mieritz, DC, M.Sc.¹, Gert Bronfort, DC, PhD^{2,4}, Markus D. Jakobsen, M.Sc.³, Per Aagaard, PhD¹, Jan Hartvigsen, DC, PhD^{1,4}

¹ University of Southern Denmark, Institute of Sports Science and Clinical Biomechanics.

² Northwestern Health Sciences University, Minneapolis, MN, USA and NIKKB.

³ National Research Centre for the Working Environment, Denmark

⁴ Nordic Institute of Chiropractic and Clinical Biomechanics, Odense, Denmark

Corresponding Author:

Rune Mygind Mieritz

Institute of Sports Science & Clinical Biomechanics, University of Southern Denmark

Campusvej 55, DK-5230 Odense M, Denmark

Ph: 0045 6550 3488

Mail: rmieritz@health.sdu.dk

Keywords: low back pain, spine, measurement, assessment, motion analysis, biomechanics, functional evaluation, instrument, device, reproducibility, reliability, agreement.

Word count abstract: 263

Word count main text: 4340

Picture: 1

Number of figures: 2

Number of tables: 4

Abstract

BACKGROUND CONTEXT: A basic premise for any instrument measuring spinal motion is that reliable outcomes can be obtained on a relevant sample under standardized conditions.

PURPOSE: The purpose of this paper was to assess the overall reliability and measurement error of regional spinal sagittal plane motion in chronic low back pain (LBP) patients. Then, to evaluate the influence of body mass index, examiner, gender, stability of pain, and pain distribution on reliability and measurement error.

STUDY DESIGN/SETTING: Test-retest design, separated by 7-14 days.

PATIENT SAMPLE: 220 chronic low back pain patients.

OUTCOME MEASURES: Kinematics of the lumbar spine were sampled during standardized spinal extension-flexion testing using a six-degree-of-freedom instrumented spatial linkage system.

METHODS: Test-retest reliability and measurement error was evaluated using Interclass Correlation Coefficients ($ICC_{1,1}$) and Bland-Altman limits of agreement (LOA).

RESULTS: The overall test-retest reliability ($ICC_{1,1}$) for various motion parameters ranged from 0.51 to 0.70 and relatively wide LOA's were observed for all parameters. Reliability measures in patient subgroups ($ICC_{1,1}$) ranged between 0.34 and 0.77. In general higher ($ICC_{1,1}$) coefficients and smaller LOA were found in subgroups with patients examined by the same examiner, patients with stable pain level, patients with BMI below 30 kg/m^2 , male patients, and patients in the Quebec Task Force classifications group 1.

CONCLUSION: This study shows that sagittal plane kinematic data from chronic low back pain patients may be sufficiently reliable in measurements of groups of patients, however because of the large LOA this test procedure appears unusable at the individual patients' level. Furthermore, reliability and measurement error varies substantially between subgroups of patients.

Introduction

When a given human kinematic quantity is repeatedly measured in the same person under standardized conditions, the outcomes typically vary between successive measurements. This may occur due to natural biological variation in the individual, variation in the measurement process, or both [1]. Quantification of this variation is crucial in order to enable clinicians to decide whether a clinically observed change represents a real change or not. One such measure of regional lumbar function is range of motion (ROM). Methods for measuring lumbar ROM include a number of technologies and methods such as inclinometers, Schober's Index, measurement of fingertip-to-floor distance, and video analysis of markers placed on anatomical landmarks. As advances in technology have occurred, devices applying computerized 3-dimensional (3D) technology have been introduced. The main advantage of these 3D instruments is their ability to provide quantitative real-time assessment of 3D regional lumbar kinematics that extends beyond the simple recording of ROM. Thus real-time information about movement velocity, acceleration and other potentially relevant parameters of the motion can be achieved, even during coupled or combined motion without any known risk to the patient. Thus, research indicates that these more advanced instruments may be useful in quantifying regional lumbar kinematics and may be valuable in generating functional diagnoses in back pain patients, while also appearing to be useful for evaluating the effectiveness of given rehabilitation therapies and for prescribing specific rehabilitation programs [2].

A potential drawback when dealing with movable instruments to record regional lumbar motion in humans is that substantial variation may originate from the measurement instrument itself as well as from the patient, the examiner, and the interface between the patient and the instrument [3;4]. Research has indicated that different measurement systems might yield non-comparable values for

the same regional lumbar movement due to differences in either the manner in which the device is attached to the participant, or the accuracy with which the device records the movements in a given plane [5;6]. However when assessing longitudinal changes in an individual's mobility using the same instrument e.g. in monitoring progress during rehabilitation, it is of primary importance to ensure that the device itself yields precise measurements and that reliable outcomes can be obtained using the same instrument.

Previous studies have described the intra- and inter-examiner reliability and measurement error of lumbar motion recording primarily in normal healthy populations [7-16] review but sources leading to variation from strict biological factors e.g. different diagnostic groups of patients with LBP, the influence of body mass index (BMI), gender, pain level, etc. have been poorly described or not addressed at all. More knowledge is needed about how these factors affect the reliability and measurement error of regional lumbar motion analysis in LBP and hence potentially limits the clinical utility of such testing in various LBP patient groups.

The current study is based on baseline trials from a randomized clinical trial [17] and therefore not specifically designed and powered for the test-retest motion analysis evaluation. This is why the reliability study may seem to have an unnecessarily high number of participants compared to studies presented in the literature [12]. However the high number of participants offered us the opportunity to stratify them into subgroups, and provide indications of factors affecting measurement error and reliability.

The overall aim of this study was to evaluate the reliability and measurement error of sagittal plane regional lumbar motion (from S1 to T7 spinous process) assessed in a large cohort of chronic LBP patients (n=220) using a non-invasive 3D measurement technology and to quantify underlying sources leading to variation between repeated measurements. Specifically, we aimed to:

1. Evaluate the overall reliability and measurement error of regional lumbar motion in the sagittal plane in 220 chronic LBP patients measured on two test occasions separated by 7 to 14 days and to evaluate the level of clinical utility (individual, group or none)
2. Evaluate the influence of BMI (greater or smaller than 30 kg m²), examiner (same or different examiner), gender (male or female), pain (stable or unstable pain level), and diagnostic group (LBP with or without radiation) on the reliability and measurement error of these measurements.

Material and Methods

Study population

Over a period of 3 years, 220 participants were recruited as part of a randomized clinical trial at the Wolfe Harris Center for Clinical Studies at Northwestern Health Science University, Minneapolis, USA [17]. Inclusion criteria were: LBP patients 18-65 years of age, Quebec Task Force classifications 1, 2, 3 and 4[18] and a primary complaint of mechanical LBP of at least 6 weeks' duration with or without radiating pain to the lower extremity. Mechanical LBP was defined as pain that had no specific identifiable etiology but that could be reproduced by back movements or provocation tests. Exclusion criteria were: previous lumbar spine fusion surgery, progressive neurological deficits, aortic or peripheral vascular disease, pain scores of less than 3 (0–10 scale), ongoing treatment for back pain by other health care providers, or participation in pending or current litigation. Participants were recruited through local newspaper advertisements, community posters, and postcard mailings and an initial screening was conducted by telephone.

Test procedures

The present test-retest design comprised two visits to the research clinic, separated by 7-14 days. The tests constituted part of the baseline examination prior to inclusion in the randomized clinical trial [17]. On the first visit, participants' anthropometrical data (height, weight) were obtained and

all participants completed a self-administered questionnaire containing information on health history and demographics. Subsequently, chiropractic and medical clinicians reviewed the health history and performed a physical examination including a complete neurological examination, orthopedic tests, and manual static and motion palpation of the lumbar spine and lower extremities. Nine trained and certified research clinicians performed the blinded, objective evaluation and outcome assessment blinded to the clinical information. Patients who qualified and agreed to participate were then scheduled for a second baseline visit in the clinic where the test procedures were repeated. LBP intensity was measured by an ordinal 11-box scale [19].

Measurement protocol

Instrumentation

Kinematics of the regional lumbar motion were sampled during a standardized extension-flexion test using a six-degree-of-freedom instrumented spatial linkage system with a sampling rate of 100 Hz (CA 6000 Spine Motion Analyzer; OSI, Union City, CA, USA). The instrument was calibrated against an inclinometer at the beginning of each test day and zero-setting was performed for each participant in the neutral position before the first test.

The CA 6000 Spine Motion Analyzer has previously been verified for precision and accuracy [5;11;14;15;20;21]. Most studies have reported the device to have good accuracy and precision, e.g. a study conducted by Feipel et al. who reported that precision was $\sim 0.3^\circ$ and accuracy $\sim 1^\circ$ [20].

Attachment and procedure

Each participant wore a loose T-shirt and trousers. The measuring device was attached with the patient standing in a neutral position with relaxed arms hanging down the side (Fig.0). The fixed extremity of the linkage was mounted on the sacral crest (S2) using the original belt. The mobile end was mounted at the level of T7 using the original chest harness and the top edge of the

horizontal metal pieces was aligned evenly with the inferior angles of the scapulae (which is level with the T7 spinous process). The pelvic harness was applied so that the binding posts were level with the posterior superior iliac spines. Neutral position was defined as the patient standing with eyes open, facing forward, with the feet positioned shoulder-width apart and arms hanging freely at the side with the low back in a comfortable position. Each patient received the following verbal explanation. *“Ok, I’ll have you find a neutral position for your low back. Place your arms across your chest and bend backward from the waist as far as you can go. As you return to neutral, move your palms to your thighs and while sliding your palms down your legs, bend forward from the waist as far as you can go, and then return to neutral (arms across chest). It should be done at your own pace and without pausing”*. Each patient then performed several trial runs as a “warm up”.

Successive trials were undertaken in each test session until reaching a ROM difference of four degrees or less between subsequent trials. A maximum of six trials were allowed, although more than two trials was rarely needed. When more than two trials were generated, a trial selection model was used to specify which trials to use (described in the data processing and analysis section below). The outcomes of the two trials were averaged and used for the analysis. In general, the two trials on the same test day were very similar. No rest period was given between trials. The time of the day was recorded but not taken into account in the planning of the two visits. The testing time duration for the complete protocol was approximately 10 minutes and included both sagittal and coronal plane motions as well as rotation and circumduction. The reliability and measurement error of motions other than those in the sagittal plane will be published in a separate publication.

Insert Fig.0 about here

Fig.0

The CA6000 Spine Motion Analyzer with the patient in a neutral (left) and flexed posture (right).

Data processing and analysis

A custom made MatLab program, produced by the authors, was used to reduce the 3Ddata into single numbered motion parameters (listed below). A trial selection model was developed in order to choose and extract the approximated two best data trials from each set. The selection process was based on a computerized comparison of all movement trials in each set.

The algorithm used to select the two best or most alike trials was performed using a pair-wise comparison based on the difference in time and distance between each trial. Each trial comparison (TC) was calculated as follows: In order to compare the vectors, regardless of movement duration, each vector was down-sampled to 100 indices. Subsequently, the total distance (D_{total}) between each pair of vectors (trial 1 vs. 2, 2 vs. 3, 1 vs. 3, etc.) was calculated (using the Euclidean distance weight function “dist”). Secondly, the difference between movement duration (T_{diff}) was divided by the average movement duration (T_{mean}).

$$TC = D_{total} + \frac{T_{diff}}{T_{mean}} \cdot 100$$

The two trials with the lowest trial comparison (TC) were then chosen as the two best trials used for statistical analysis.

In addition, a manual visual inspection of all sagittal motion plots was done in conjunction with the computerized selection process.

The following parameters were selected on the basis of the current study hypothesis, clinical experience and previous study findings [2;22;23]:

1. *ROM (degree)* was calculated as the total angular range of regional lumbar motion in the sagittal plane from maximum extension to maximum flexion (Figure 1).
2. *Mean flexion velocity (degree/sec)* was calculated using the central limit theorem as the average angular speed from maximum extension to maximum flexion position (Figure 1).
3. Maximum flexion and extension velocity: *Maximum flexion velocity (degree/second)* was calculated as the peak angular speed in the forward bending motion reached from full extension to full flexion. *Maximum extension velocity (degree/second)* was calculated as the peak negative angular speed reached in the ROM from full flexion and back to the neutral position (Figure 1).
4. Phase-plot area ($\text{degree}^2/\text{sec}$) was defined as the area comprised by the phase-plot of sagittal flexion-extension angular motion versus velocity (Figure 2). Phase-plot area was calculated based on cross-product calculations between vectors drawn from the neutral position (0,0) to each coordinate point in the phase-plot. Motion phase (position-velocity) plot area previously has been used to quantify human motion characteristics [24;25]. However, to our best knowledge, motion phase area has not been used in the quantification of regional lumbar movements. Notably, phase plot area is sensitive to changes in both sagittal ROM and velocity, while accounting for the cumulated time-history of motion throughout the entire ROM (Figure 2).
5. *Jerk Index* was calculated from maximum extension to maximum flexion as the mean spectral frequency of the first derivative of the angular acceleration signal multiplied by movement duration. This parameter indicates the number of changes in acceleration, i.e. the smoothness of the motion.

$$Jerk\ index = mpf \left(\frac{dA}{dt} \right) * t$$

A = acceleration, mpf = mean power frequency, t = duration of the movement

Insert Fig.1 about here

Fig.1

Regional lumbar extension and flexion motion displayed in degrees and velocity at a single representative session. *Note: Average velocity was measured as the slope of the straight line between maximum extension and flexion positions.*

Insert Fig.2 about here

Fig.2

Regional lumbar extension and flexion motion obtained in a single representative participant displayed as a position (degree)-speed (velocity) phase plot. The area within the two curves is defined as the regional lumbar motion area. *Note: Movement was started at (0,0) from which the participant performed an extension motion to reach maximum extension ROM (negative position and velocity values) followed by a flexion motion to reach maximum flexion ROM (positive velocity values) and subsequently returned to neutral position (negative velocity values, positive position values).*

Statistical analysis

The concepts of reliability and measurement error were used based on the COSMIN study taxonomy, terminology and definitions [26]. The measurement properties of reliability and measurement error were defined in the following way: reliability: the proportion of the total variance in the measurements which is because of ‘true’ differences among patients; and

measurement error: the systematic and random error of a patient's score that is not attributed to true changes in the construct [26].

A number of statistical tools were used to assess test-retest reliability and measurement error. Paired t-testing was used to detect systematic bias between tests sessions. Based on the study design i.e. each target was rated by a different mix of the nine examiners (considered to be randomly selected from a larger population of assessors) and because we aimed to generalize to individual ratings, the Intraclass Correlation Coefficients ($ICC_{1,1}$) were calculated to assess reliability [27]. To assess measurement error, Bland-Altman limits of agreement (LOA) with 95% confidence intervals (CI) were calculated [28].

The assumption of normally distributed and homoscedastic data was tested based on Bland-Altman plots and correlation analysis between absolute differences and mean values. In addition, the Shapiro-Wilk W test was performed for the residuals in order to check for normal distribution. When heteroscedastic relationships were found, a natural log transformation was applied prior to statistical analysis.

The stratification into subgroups according to BMI was based on the cut-off points proposed by the WHO for the classification of overweight i.e. a BMI greater or equal to 30 [29]. The subgroup including patients with an unstable pain level was defined as patients with a change in pain intensity score between test and retest of ± 2 or more [30]. Diagnostic group stratification was done so that diagnostic groups 2, 3 and 4 were collapsed, dividing patients into two groups including LBP with or without radiation.

Stata software version 11 was used for all statistical analyses.

Results

A total of 220 LBP patients participated in the study. All participants suffered from chronic back pain and functional disability (Table 1). Initially, 301 patients were recruited but due to technical problems, 220 full patient recordings were obtained. The individuals not available for analysis were slightly younger i.e. having a median age of 43 years and 95% CI ranging from 40 – 47 years vs. a median of age of 47 years and 95% CI (45 – 49 years) in the 220 patients included in this study. There was no difference in other baseline characteristics such as BMI, gender, duration of pain, or depression score (CESD), LBP intensity and leg pain intensity (ordinal 11-box scale).

From the clinical findings classified by signs and symptoms, the LBP patients were fitted into the following diagnostic groups using the Quebec Task Force classification system:

- Group 1: LBP patients without radiation (n=149)
- Group 2: LBP patients with pain radiation to proximal lower extremity (n=40)
- Group 3: LBP patients with pain radiation to distal parts of lower extremity (n=27)
- Group 4: LBP patients with pain radiation to lower extremity accompanied by neurological signs (n=4)

Insert Table 1 about here

A statistical difference was observed between Session 1 and Session 2 for the three motion parameters: mean flexion velocity (16% higher, $p=0.001$), maximal flexion velocity (6% higher $p=0.011$) and Jerk Index (22% lower, $p=0.001$), (Table 2). All other parameters showed no systematic variation between Sessions 1 and 2.

Insert Table 2 about here

All regional lumbar motion parameters except total ROM showed a heteroscedastic appearance when displayed in Bland-Altman plots, implying that the magnitude of variance (error) was associated with the magnitude of the parameter [28]. When these measurements were log transformed, the mean and SD of the test-retest differences remained more similar throughout the range of parameter values, indicating that this procedure would provide a better fit with the statistical model. The Shapiro-Wilk W test calculated on the residuals indicated a more normal distribution of the data when log transformation was performed except for the ROM parameter. Therefore, $ICC_{(1,1)}$ and Bland-Altman LOA were calculated with logarithmically transformed data, except for total ROM [28]. $ICC_{(1,1)}$ values for the motion parameters calculated on the basis of the total LBP group ranged between 0.51 and 0.70. Relatively wide LOA values were observed for all parameters (Table 3).

Insert Table 3 about here

The large number of participants examined allowed us to analyze specific subgroups. Reliability data from the subgroup analysis ($ICC_{(1,1)}$) ranged between 0.34 and 0.77. The subgroup analysis revealed that LBP patients with an unstable pain level (ordinal 11-box scale change ± 2 or more) had a larger variation between test sessions compared to participants with stable pain. The largest variation between the pain-stable and pain-unstable subgroups was found in the motion parameter maximum flexion velocity ($ICC_{(1,1)}$ 0.69 vs. 0.48, LOA_{ratio} 0.51 to 1.65 vs. 0.47 to 2.19).

Furthermore, a general trend was that higher $ICC_{(1,1)}$ coefficients and smaller LOA values were found in patients examined by the same examiner, patients with BMI below 30, patients with male

gender, and patients in the Quebec Task Force classification group 1. The results from the subgroup analysis are shown in Table 4.

Insert Table 4 about here

Discussion

This is, to our knowledge, the largest reliability study dealing with regional lumbar motion in chronic LBP patients and the first one to analyze subgroups. This study shows that regional lumbar motion data from chronic LBP patients may be sufficiently reliable to be used for statistical analysis at a group level but it is unlikely that it can be used at the individual patient level. The Jerk Index however may not be suitable for analysis at either group or individual level. In addition, we found that reliability and measurement error vary between subgroups of patients.

A systematic difference between test and retest sessions was found for mean and maximal flexion velocity and Jerk Index (cf. Table 2) indicating the presence of a learning effect caused by patients' habituation to the instrument and surroundings. Notably however, mean extension velocity as well as total ROM and position-velocity phase-plot area did not show any systematic difference between the two sessions.

In terms of reliability measures, the ICC provides a unitless index based on the ratio of the within-to-between subject test-retest differences and ranges from 0.00 to 1.00, with values closer to 1.00 representing stronger reliability. There are no standard values for acceptable reliability and ICC values reported in the literature are therefore difficult to compare. Furthermore, ICCs are not only determined by measurement error but also by between-subject variation and ICCs can be calculated differently, yielding different results [27;31]. As a guideline, Portney and Watkins suggest that values below 0.50 represent poor reliability, coefficients from 0.50 to 0.75 suggest moderate

reliability, and values above 0.75 are indicative of good or high reliability [32]. Aaronson et al. recommend that minimal standards for reproducibility coefficients to be 0.70 for group comparisons and 0.90–0.95 for individual measurements over time [33].

Our subgroup analysis indicated that certain factors affect the level of reliability and measurement error more than others. Thus, subgroup ICCs_(1,1) ranged from 0.77 when total ROM was assessed in subgroups by the same examiner to 0.34 for the Jerk Index in the subgroup of females. Highest ICC_(1,1) and smallest LOA were found in the groups examined by the same examiner, with a BMI below 30, a stable pain level between tests, of male gender, and in the Quebec Task Force classification group 1. This information indicates that regional lumbar motion parameters provide useful information provided that these parameters are controlled.

To evaluate changes over time in an individual, the change must exceed the inherent variability of the measurements, which are determined using the LOA. In this study, the same examiner subgroup showed the narrowest LOA_{degree} for total ROM, ranging from -20 to 22 degrees. Maximum extension velocity for participants with a BMI lower than 30 showed the narrowest LOA ratio (0.58 to 1.60) i.e. in a given LBP patient and for this test parameter, one test result may differ from another test result within this ratio. The Jerk Index for participants with an unstable pain level showed the widest LOA_{ratio}, ranging from 0.48 to 2.95. Collectively, LOA intervals were all relatively wide indicating that a quite substantial change would be required for a given individual patient to confidently state that an actual change had taken place in that individual.

Previous biomechanical studies using similar technology in test-retest studies have focused mainly on measurements of ROM in healthy individuals without LBP. Using the OSI CA 6000, high intra- and inter-examiner reliability have been demonstrated for lumbar ROM in asymptomatic individuals [8;9;11;14]. Schuit et al. examined ROM in LBP patients using repeated trials on the same day 2 minutes apart and reported intra-examiner reliability ICC_(2,1) values that ranged from

0.875 to 0.966 and standard error of measurement (SEM) in regional lumbar flexion of 3.7° and in extension of 2.8°. Also, these authors suggested an acceptable validity for this method when compared to X-ray analysis [15]. In order to compare their SEM values to our results, we can estimate the smallest detectable change (SDC) for a 95% confidence level in flexion as $SEM \cdot \sqrt{2} \cdot 1.96$, where SEM is the square root of the error variance[31]. This results in an SDC for flexion of 10.3° and an SDC for extension of 7.7°. When measurements in a given individual differ by more than these values, it can be concluded that the difference represents a real change. Although different study results cannot be compared directly because of differences in experimental designs and statistical analyses, the current results indicate that the observations by Schuit et al. may be difficult to replicate in LBP patients when successive tests are separated by 7-14 days (current study) instead of 2 minutes [15]. When test-retest was performed by the same examiner, we found that $ICC_{(1,1)}$ for total ROM was 0.77 while 95% LOA ranged from -20° to 22°. Thus, the current results indicate that LBP patients have a considerable natural variation in regional lumbar motion when evaluated in successive test sessions separated by days to weeks.

In clinics, measurement error is very important because patients are managed at the individual level. Measurement error cannot be reduced by the sample size as in research, only by repetition of measurements, which is often too time-consuming in daily clinical practice. Therefore, clinicians are advised to have a continuous focus on standardizing objective measurement procedures as much as possible in order to reduce measurement error. This study indicates that this technology is not yet at the stage of development that it can be recommended for use in a clinical setting.

In research, comprehensive and transparent reporting of all known and plausible factors potentially affecting measurement error and reliability are fundamental in future test-retest studies. Key areas for future studies are likely to include detailed descriptions of the standardization of measurement procedures and instructions to examiners. The equipment used for kinematic assessment is

generally very precise - the challenges appear when humans get involved. Therefore, procedures and attachment techniques should be further developed in order to reduce measurement bias. Future investigations should perform test-retest studies on all kinds of motion parameters, not only ROM.

Study Limitations

Several potential methodological limitations were noted in the current study. The trial selection process for all patients on each test-day was based on a computerized comparison with an additional visual inspection of all sagittal plane plots. The inspection process was limited by requiring a subjective judgment but this process made it possible to identify potential errors not detected by the computerized comparison. Any subjective judgment or manual override was however rarely needed (less than 10 occasions).

The fixed extremity of the CA6000 Spine Motion Analyzer was mounted on the sacral crest (S2) and the mobile end of the device was mounted at the level of T7 spinous process. Therefore the measured movements were due both lumbar motion and motion from the lower part of the thoracic region.

To reduce complexity and data abundance, the recorded regional lumbar motion data were reduced into a number of separate motion parameters. The selected parameters could be considered as reductionist models to achieve descriptive measures of complex lumbar movement patterns in LBP patients at the functional level. Obviously, the choice of parameters and analysis approach means that other types of important kinematic information may remain uncovered. In addition, these results cannot be extrapolated to LBP patients. Notably therefore, the current results should not be generalized to other populations since fluctuating pain conditions are likely to have influenced the results.

Conclusion

Most sagittal plane regional lumbar motion parameters obtained in pain-stable chronic low back pain patients indicate that motion data recorded using this setup could be used for group comparison but not to assess individual patients. The limits of agreement data indicate that minor-to-moderate changes in individual patients cannot be detected with the current measurement setup, i.e. a relatively large change is required in order to be 95% confident that a real change has taken place in a given patient.

In addition this study shows that reliability and measurement error of regional motion parameters varies substantially between subgroups of patients.

References

- [1] Bland JM, Altman DG. Statistics notes: Measurement error. *BMJ* 1996;312:1654.
- [2] Lehman GJ. Biomechanical assessments of lumbar spinal function. How low back pain sufferers differ from normals. Implications for outcome measures research. Part I: kinematic assessments of lumbar function. *J Manipulative Physiol Ther* 2004;27:57-62.
- [3] Morl F, Blickhan R. Three-dimensional relation of skin markers to lumbar vertebrae of healthy subjects in different postures measured by open MRI. *Eur Spine J* 2006;15:742-51.
- [4] Mayer TG, Kondraske G, Beals SB, Gatchel RJ. Spinal range of motion - Accuracy and sources of error with inclinometric measurement. *Spine* 1997;22:1976-84.
- [5] Christensen HW. Precision and accuracy of an electrogoniometer. *J Manipulative Physiol Ther* 1999;22:10-4.
- [6] Mannion A, Troke M. A comparison of two motion analysis devices used in the measurement of lumbar spinal mobility. *Clin Biomech* 1999;14:612-9.
- [7] Barrett CJ, Singer KP, Day R. Assessment of combined movements of the lumbar spine in asymptomatic and low back pain subjects using a three-dimensional electromagnetic tracking system. *Man Ther* 1999;4:94-9.

- [8] Dopf CA, Mandel SS, Geiger DF, Mayer PJ. Analysis of spine motion variability using a computerized goniometer compared to physical examination. A prospective clinical study. *Spine* 1994;19:586-95.
- [9] Dvorak J, Vajda EG, Grob D, Panjabi MM. Normal motion of the lumbar spine as related to age and gender. *Eur Spine J* 1995;4:18-23.
- [10] Gill KP, Callaghan MJ. Intratester and intertester reproducibility of the lumbar motion monitor as a measure of range, velocity and acceleration of the thoracolumbar spine. *Clin Biomech* 1996;11:418-21.
- [11] McGregor AH, McCarthy ID, Hughes SP. Motion characteristics of the lumbar spine in the normal population. *Spine* 1995;20:2421-8.
- [12] Mieritz RM, Bronfort G, Kawchuk G, Breen A., Hartvigsen J. Reliability and measurement error of 3D regional lumbar motion measures: a systematic review. *J Manipulative Physiol Ther* 2012; Accepted for publication.
- [13] Moutzouri M, Billis E, Strimpakos N, Kottika P, Oldham JA. The effects of the Mulligan Sustained Natural Apophyseal Glide (SNAG) mobilisation in the lumbar flexion range of asymptomatic subjects as measured by the Zebris CMS20 3-D motion analysis system. *BMC Musculoskel Dis* 2008;9:131.
- [14] Petersen CM, Johnson RD, Schuit D, Hayes KW. Intraobserver and interobserver reliability of asymptomatic subjects' thoracolumbar range of motion using the OSI CA 6000 Spine Motion Analyzer. *J Orthop Sports Phys Ther* 1994;20:207-12.
- [15] Schuit D, Petersen C, Johnson R, Levine P, Knecht H, Goldberg D. Validity and reliability of measures obtained from the OSI CA-6000 Spine Motion Analyzer for lumbar spinal motion. *Man Ther* 1997;2:206-15.
- [16] Troke M, Moore AP, Cheek E. Intra-operator and inter-operator reliability of the OSI CA 6000 Spine Motion Analyzer with a new skin fixation system. *Man Ther* 1996;1:92-8.
- [17] Bronfort G, Maiers MJ, Evans RL et al. Supervised exercise, spinal manipulation, and home exercise for chronic low back pain: a randomized clinical trial. *Spine J* 2011;11:585-98.
- [18] Atlas SJ, Deyo RA, Patrick DL, Convery K, Keller RB, Singer DE. The Quebec Task Force classification for spinal disorders and the severity, treatment, and outcomes of sciatica and lumbar spinal stenosis. *Spine* 1996;21:2885-92.
- [19] Jensen MP, Karoly P, Braver S. The measurement of clinical pain intensity: a comparison of six methods. *Pain* 1986;27:117-26.
- [20] Feipel V, Rondelet B, Le Pallec JP, Rooze M. Normal global motion of the cervical spine: an electrogoniometric study. *Clin Biomech* 1999;14:462-70.
- [21] Dvorak J, Antinnes JA, Panjabi M, Loustalot D, Bonomo M. Age and gender related normal motion of the cervical spine. *Spine* 1992;17:S393-S398.

- [22] Jordan K, Dziedzic K, Mullis R, Dawes PT, Jones PW. The development of three-dimensional range of motion measurement systems for clinical practice. *Rheumatology* 2001;40:1081-4.
- [23] Sjolander P, Michaelson P, Jaric S, Djupsjobacka M. Sensorimotor disturbances in chronic neck pain - Range of motion, peak velocity, smoothness of movement, and repositioning acuity. *Man Ther* 2008;13:122-31.
- [24] Gullam JE, Blanks AM, Thornton S, Shmygol A. Phase-plot analysis of the oxytocin effect on human myometrial contractility. *Eur J Obstet Gynecol Reprod Biol* 2009;144 Suppl 1:S20-S24.
- [25] Prados F, Boada I, Prats-Galino A et al. Analysis of new diffusion tensor imaging anisotropy measures in the three-phase plot. *J Magn Reson Imaging* 2010;31:1435-44.
- [26] Mokkink LB, Terwee CB, Patrick DL et al. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *J Clin Epidemiol* 2010;63:737-45.
- [27] Shrout PE, Fleiss JL. Intraclass Correlations - Uses in Assessing Rater Reliability. *Psychol Bull* 1979;86:420-8.
- [28] Bland JM, Altman DG. Measuring agreement in method comparison studies. *Stat Methods in Med Res* 1999;8:135-60.
- [29] WHO Expert Committee. Physical status: The use and interpretation of anthropometry. Geneva: WHO; 1995.
- [30] Dworkin RH, Turk DC, Wyrwich KW et al. Interpreting the clinical importance of treatment outcomes in chronic pain clinical trials: IMMPACT recommendations. *J Pain* 2008;9:105-21.
- [31] De Vet HCW, Terwee CB, Knol DL, Bouter LM. When to use agreement versus reliability measures. *J Clin Epidemiol* 2006;59:1033-9.
- [32] Portney LG, Watkins M.P. *Foundation of clinical research: applications to practice*. Third ed. Upper Saddle River, New Jersey: Pearson Education, Inc.; 2009.
- [33] Aaronson N, Alonso J, Burnam A et al. Assessing health status and quality-of-life instruments: Attributes and review criteria. *Qual Life Res* 2002;11:193-205.



Fig. 0: The CA6000 Spine Motion Analyzer with a person in neutral, extension and flexion positions

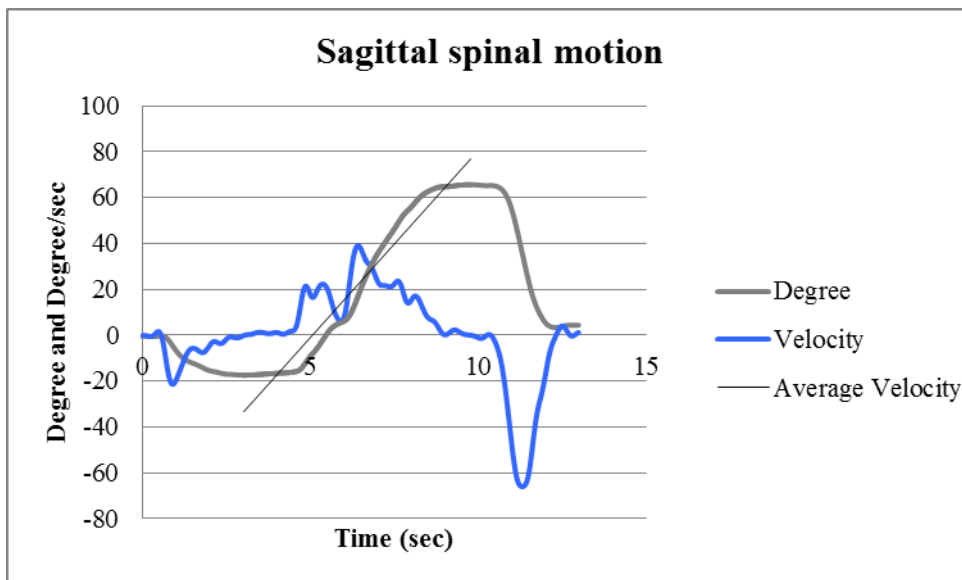


Fig.1

Regional lumbar extension and flexion motion displayed in degrees and velocity at a single representative session. *Note: Average velocity was measured as the slope of the straight line between maximum extension and flexion positions.*

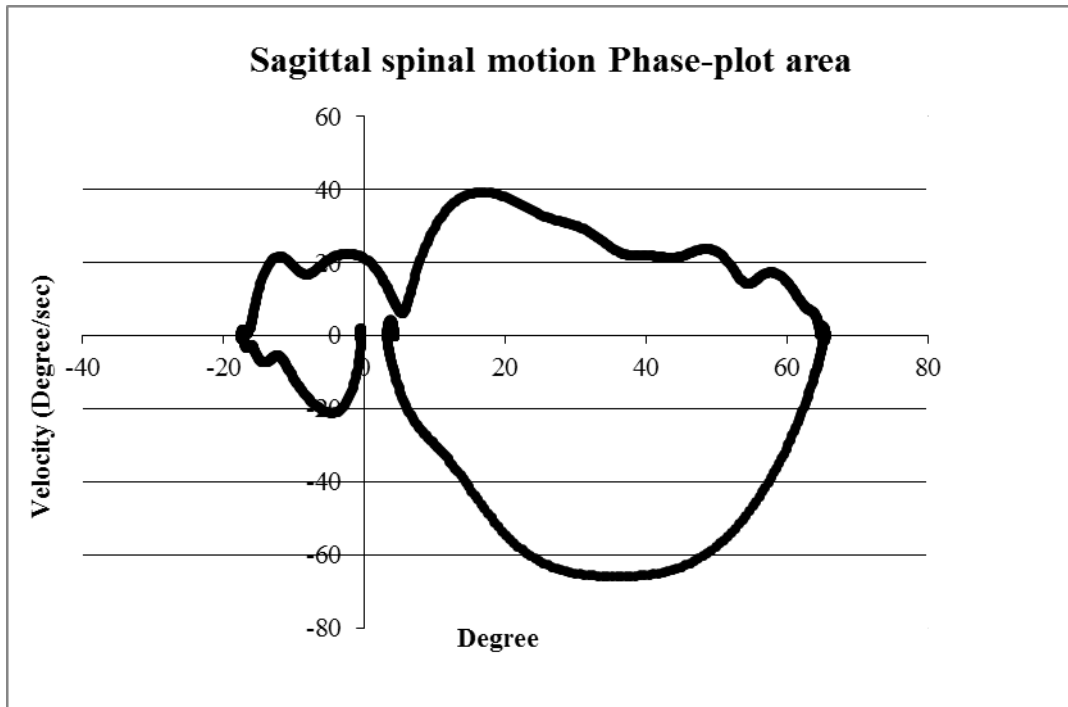


Fig.2

Regional lumbar extension and flexion motion obtained in a single representative participant displayed as a position (degree)-speed (velocity) phase plot. The area within the two curves is defined as the regional lumbar motion area. *Note: Movement was started at (0,0) from which the participant performed an extension motion to reach maximum extension ROM (negative position and velocity values) followed by a flexion motion to reach maximum flexion ROM (positive velocity values) and subsequently returned to neutral position (negative velocity values, positive position values).*

Table 1

Subject characteristics for the present group of male and female LBP patients

	Male (n = 87)		Females (n = 133)	
	Mean (SD)	Range	Mean (SD)	Range
Age (years)	44.7 (11.3)	22-64	46.0 (11)	22-65
Height (cm)	178 (6.3)	164-196	163.3 (8.3)	133-183
Weight (kg)	89.2 (15.2)	66-159	75.1 (17)	45-133
BMI	28.2 (6.2)	18.7-47.2	28.1(4.7)	19.3-49.6
Activity level	2.4 (1.3)	0-4	2.4 (1.2)	0-4
Past episodes of back pain	3.4 (0.6)	1-4	3.2(0.9)	0-4
S1 LBP past week	5.1 (1.6)	3-9	5.2 (1.6)	3-10
S2 LBP past week	5.3 (1.6)	2-8	5.1 (1.8)	2-10
S1 LP past week	2.1 (2.4)	0-9	2.3 (2.6)	0-10
S2 LP past week	2.0 (2.5)	0-8	2.1 (2.5)	0-10
S1 Roland Morris (0-23)	8.1 (4.4)	2-17	8.8 (4.3)	1-21
S2 Roland Morris (0-23)	8.0 (5.0)	0-23	8.8 (4.9)	0-21

Activity level: *Engaged in exercise or sports activities in the past month? (0 = I do not engage in exercise or sports, 1 = Less than once a week, 2 = Once a week, 3 = 2 or 3 times per week, 4 = 4 times or more per week)*
 Past episodes of LBP: *An episode is a week with at least some LBP (0 =None, 1 = 1-2 episodes, 2 = 3-5 episodes 3=More than 5 episodes, 4= 1 single episode of continuous LBP)*
 (S) =session, (LP)= leg pain, (LBP) =low back pain (measured on a numerical pain scale “one to ten”, ten being the worst possible pain).

Table 2

Motion parameters recorded during voluntary lumbar sagittal plan motion

Parameter	Session 1			Session 2			p
	Mean (SD)	Min	Max	Mean (SD)	Min	Max	
ROM (°)	71 (16)	31	111	69 (16)	26	113	0.060
Flexion mean velocity (°/sec)	11.5 (5.7)	2.2	41.6	13.3 (6.2)	3.9	40.9	0.001
Extension max velocity (°/sec)	-37.2 (14.9)	-93	-12	-37.8 (12.8)	-75	-10	0.123
Flexion max velocity (°/sec)	34.3 (13.4)	10.3	93.7	36.2 (13.7)	10.8	98.7	0.011
Phase-plot Area (°*(°/sec))	4020 (2040)	926	15923	3994 (1797)	926	10135	0.998
Jerk index	12.7 (5.7)	4	34	10.4 (4.5)	2	29	0.001

Means, standard deviations, range and P-value for all 220 patients flexion motion measurements with the OSI Spine Motion Analyzer.

Table 3

Reliability and measurement error of lumbar spinal motion parameters in the sagittal plane

Parameter (n=220)	ICC _(1,1)	(95% CI)		95% LOA		
ROM (°)	0.69	0.62	0.76	-23	-	27
Flexion mean velocity (°/sec)	0.61	0.53	0.70	0.40	-	1.83
Extension max velocity (°/sec)	0.70	0.63	0.76	0.55	-	1.71
Flexion max velocity (°/sec)	0.64	0.56	0.72	0.50	-	1.79
Phase-plot Area (°*(°/sec))	0.69	0.62	0.76	0.47	-	2.11
Jerk index	0.51	0.42	0.61	0.57	-	2.59

ICC_(1,1) = Intraclass correlation coefficient, LOA = Limits of agreement, ROM = Range of motion, Jerk Index = number of changes in acceleration from full extension to full flexion, *All parameters except ROM was natural log transformed to fit the statistical model and is therefore presented in LOA(ratio).

Table 4

Spinal motion reliability and measurement error for LBP patients divided into subgroups

Motion Parameter	Statistical parameter	Subgroups									
		Same ex.	Different ex.	BMI <30	BMI >30	Pain (s)	Pain (u)	Male	Female	Group 1	Group 2,3,4
	Number of subjects	89	131	147	73	162	58	87	133	149	71
ROM	ICC _(1,1)	0.77	0.64	0.66	0.73	0.69	0.7	0.66	0.65	0.68	0.69
	LOA _{UL} (degree)	22	30	26	27	26	29	24	28	26	28
	LOA _{LL} (degree)	-20	-25	-24	-21	-22	-28	-25	-22	-25	-21
Mean velocity	ICC _(1,1)	0.54	0.65	0.63	0.58	0.64	0.53	0.7	0.53	0.65	0.55
	LOA _{UL} (ratio)	1.79	1.86	1.8	1.89	1.71	2.18	1.66	1.95	1.71	2.1
	LOA _{LL} (ratio)	0.39	0.4	0.42	0.36	0.41	0.37	0.42	0.39	0.41	0.38
Max flexion velocity	ICC _(1,1)	0.67	0.61	0.67	0.58	0.69	0.48	0.72	0.57	0.66	0.59
	LOA _{UL} (ratio)	1.6	1.91	1.68	2.03	1.65	2.19	1.72	1.85	1.69	2.02
	LOA _{LL} (ratio)	0.49	0.51	0.51	0.48	0.51	0.47	0.52	0.48	0.51	0.48
Max extension velocity	ICC _(1,1)	0.74	0.67	0.75	0.6	0.71	0.63	0.71	0.68	0.69	0.7
	LOA _{UL} (ratio)	1.63	1.77	1.6	1.93	1.67	1.83	1.69	1.73	1.68	1.79
	LOA _{LL} (ratio)	0.57	0.53	0.58	0.5	0.55	0.55	0.53	0.57	0.56	0.53
Phase-plot Area	ICC _(1,1)	0.76	0.64	0.7	0.68	0.74	0.56	0.7	0.66	0.7	0.67
	LOA _{UL} (ratio)	1.87	2.27	1.95	2.41	1.92	2.65	1.87	2.26	1.92	2.5
	LOA _{LL} (ratio)	0.5	0.46	0.48	0.47	0.5	0.42	0.48	0.47	0.49	0.45
Jerk Index	ICC _(1,1)	0.50	0.52	0.56	0.41	0.55	0.42	0.70	0.34	0.55	0.41
	LOA _{UL} (ratio)	2.60	2.59	2.45	2.86	2.45	2.95	2.31	2.78	2.53	2.72
	LOA _{LL} (ratio)	0.58	0.56	0.58	0.56	0.61	0.48	0.61	0.55	0.57	0.57

Ex. = examiner(s), BMI = body mass index (kg/m²), Pain (s) = pain score max change ± 1 , Pain (u) = pain score change ± 2 or more, Group = Quebec Task Force classifications 1 vs. 2, 3 and 4, ROM = Range of motion (degree), Velocity = (degree/sec), Phase-plot = Phase-plot Area (degree²/sec), Jerk Index = number of changes in acceleration from full extension to full flexion, ICC_(1,1) = Intraclass correlation coefficient, LOA_{UL} = Limits of agreement (upper limit), LOA_{LL} = Limits of agreement (lower limit), NB: Max and min values in each row are bolded.

Manuscript III

Regional lumbar motion changes in chronic low back pain patients

A secondary analysis of data from a randomized clinical trial

Rune M. Mieritz, DC, M.Sc.¹, Jan Hartvigsen, DC, PhD^{1,2}, Eleanor Boyle, PhD^{1,3}, Markus D. Jakobsen, M.Sc.^{1,4}, Per Aagaard, PhD¹, Gert Bronfort, DC, PhD^{2,5}

¹Institute of Sports Science and Clinical Biomechanics, University of Southern Denmark, Odense, Denmark; ²Nordic Institute of Chiropractic and Clinical Biomechanics, Odense, Denmark; ³Dalla Lana School of Public Health, University of Toronto, Toronto, Canada; Division of Healthcare and Outcomes Research, Toronto Western Research Institute, University Health Network, Toronto, Canada; ⁴National Research Centre for the Working Environment, Copenhagen, Denmark; ⁵Northwestern Health Sciences University, Minneapolis, USA .

Corresponding Author:

Rune Mygind Mieritz

Institute of Sports Science and Clinical Biomechanics, University of Southern Denmark

Campusvej 55, DK-5230 Odense M, Denmark

Phone: 0045 61 308 589

E-mail: rmieritz@health.sdu.dk

Keywords: low back pain, spine, measurement, assessment, motion analysis, biomechanics, nonsurgical, treatment, manipulation, exercise.

Word count abstract: 269

Word count main text: 4384

Number of figures: 5

Number of tables: 2

Abstract

BACKGROUND CONTEXT: Several therapies have been used in the treatment of chronic low back pain, including various exercise strategies and spinal manipulative therapy. A common belief is that spinal motion changes in particular ways in direct response to specific interventions, such as exercise or spinal manipulation.

PURPOSE: The purpose of this study was to assess changes in regional lumbar motion over 12 weeks by evaluating four motion parameters in the sagittal plane and two in the horizontal plane in LBP patients treated with either exercise therapy or spinal manipulation.

STUDY DESIGN/SETTING: Secondary analysis of a subset of participants from a randomized clinical trial.

PATIENT SAMPLE: 199 study participants with low back pain of more than six weeks' duration who had regional lumbar motion measures obtained before and after the period of intervention.

OUTCOME MEASURES: Regional lumbar motion data were sampled using a six-degrees-of-freedom instrumented spatial linkage system.

METHODS: Trained therapists collected regional lumbar motion data at baseline and at 12 weeks follow up. The regional lumbar motion data were analyzed relative to treatment modality (high-dose, supervised low-tech trunk exercise, spinal manipulative therapy, and a short course of home exercise and self-care advice).

RESULTS: The group receiving spinal manipulation changed significantly in all, and the exercise groups in half, the motion parameters included in the analysis. The spinal manipulation group changed to a smoother motion pattern (reduced Jerk Index) while the exercise groups did not.

CONCLUSION: This study provides evidence that regional lumbar motion changes can occur in chronic low back pain patients over a 12-week period and that these changes are associated with the type of treatment received.

Introduction

Chronic low back pain is a common and disabling condition for individuals and a substantial economic burden for societies [1;2]. For many years, researchers and clinicians have sought to measure back problems objectively, primarily to attempt to determine the origin of pain, and subsequently, to measure whether given types of treatment evoke a biologically or biomechanically measurable change. There is a tradition of basing diagnoses on the results of imaging techniques, such as conventional X-ray, CT or MRI [3;4]. Ascribing a patient's low back pain (LBP) to a presumed injured or painful structure (i.e., a pathoanatomical source) is often inaccurate even when based on advanced imaging techniques [5;6]. In many cases, LBP patients may show no identifiable pathoanatomical source [5-7]. Conversely, it is not a rare observation that asymptomatic individuals demonstrate spinal pathologies evident on imaging [7;8]. Consequently, it has been proposed that spinal physical impairment and disability are better evaluated by assessing measurements of the movement pattern in specific motor tasks and/or recording of maximal muscle strength/power to determine the patient's functional ability [9;10]. Functional capacity assessments addressing strength and endurance of trunk musculature can be performed to monitor the problem of LBP impairment. However, they are limited in that they measure 'extreme' capacity, which often goes beyond 'normal' trunk function needed for typical activities of daily living (ADL) [10].

Traditionally in the clinic, regional lumbar movement is quantified by measuring e.g. range of motion (ROM) or Schober's index [11]. Such 'low tech' measurements describe the full functional range of joint excursion but little about the quality of the motion. Research has indicated that simple ROM measurements have limited use as a measure of treatment outcomes or as a stand-alone measure of disability [12;13]. It has been proposed that a link between lumbar motion and lumbar pain may be found by addressing the patterns or higher order kinematics of the motion rather than the end ranges of motion [9;14;15]. More advanced motion parameters derived from 'high tech'

computerized real-time motion devices may contribute to describing patient movement and movement changes [9;15].

Several motion parameters can be derived from real-time regional lumbar motion analysis e.g. angular velocity, acceleration and smoothness of motion, respectively [16]. The development of advanced techniques to measure trunk motion characteristics during unloaded free dynamic activities represents an attempt to remedy existing deficiencies in the quantification of LBP impairment [9;15;17;18].

Many hypotheses and theories exist about how different treatment modalities such as exercise or spinal manipulation affect biomechanical spine function [19;20]. Several specific therapies have demonstrated positive effect on patient-reported outcomes [21-24] but little is known about the change in regional lumbar movement characteristics following treatment. When a therapist treats a patient, a common belief is that regional lumbar motion changes in particular ways in direct response to specific interventions, such as exercises or manual therapy. However, there seems to be a lack of science-based knowledge on this important aspect of clinical rehabilitation.

McGregor et al. reported that acute LBP patients exhibited a ‘stepped’ flexion extension motion and moved slower [15]. ROM may be the most common motion parameter [25] but velocity has been proposed to be a more sensitive parameter [15]; however these may not capture the motion patterns. Therefore we have developed new motion parameters that might better reflect the patient’s motion patterns in terms of smoothness of motion (Jerk index) and functional ability.

The overall aim of the current study was to analyze changes in regional lumbar (from S1 to T7 spinous process) motion over 12 weeks by describing pre-post treatment changes in the different treatment groups: spinal manipulation therapy (SMT), high dose supervised trunk exercise (SET), and low dose home exercise and advice (HEA), by evaluating four motion parameters in the sagittal plane (ROM, maximum flexion velocity, phase-plot area, smoothness of motion [Jerk Index]) and

two circumduction area motion parameters in the horizontal plane. The HEA group was intentionally minimal in its approach so it could serve as a control [22].

Specifically, we wanted to quantify the effect of 12 weeks of SMT, SET or HEA on regional lumbar motion in chronic LBP patients. Based on the assumption that the amount of physical modification is related to the amount of change in regional lumbar motion, we hypothesized that the groups receiving either SET or SMT care would change significantly in all motion parameters over a 12-week period whereas the minimal intervention group (HEA) would not (no change in motion parameters). Specifically, the Jerk Index was hypothesized to decrease and all other motion parameters to increase.

Material and Methods

Design

This regional lumbar motion analysis study is a secondary analysis of a subset of study participants from an observer-blinded, parallel-group, randomized clinical trial (RCT) [22]. Thus the current study is based on a secondary aim, which was to evaluate if there are treatment group differences in objective lumbar spine function after 12 weeks of treatment.

Participants were consecutively recruited over a period of 3 years at the Wolfe Harris Center for Clinical Studies at Northwestern Health Sciences University, Minneapolis, USA. The institutional review boards of the Northwestern Health Sciences University, the Minneapolis Medical Research Foundation, and the University of Minnesota approved the study and written informed consent was obtained from all study participants. Regional lumbar motion recordings were measured on two baseline (PRE) visits (separated by 7-14 days) and one follow-up visit after 12 weeks of intervention (POST). To illustrate the stability of pain intensity in the overall cohort, pain intensity levels recorded at the two baseline measurement time points are presented in Table 1.

Participants

Inclusion/exclusion criteria

For the current study, the participants must have been randomized in the RCT and have completed baseline and follow up regional lumbar motion assessment procedures. The RCT had the following inclusion criteria: patients suffering from pain, stiffness, or tenderness in the low back between the lower margin of the ribs and the gluteal fold, with or without concurrent leg pain and/or neurological signs of more than 6 weeks' duration, be 18-65 years of age, and categorized as Quebec Task Force classifications 1 (pain without radiation), 2 (pain with proximal extremity radiation), 3 (pain with distal extremity radiation) or 4 (pain with radiation and neurologic finding) [26]. Exclusion criteria were: previous lumbar spine surgery, back pain referred from local joint lesions of the lower extremities or from known visceral disease, progressive neurological deficits due to nerve root or spinal cord compression, aortic and peripheral vascular disease, cardiac disease requiring medical treatment, blood clotting disorders, diffuse idiopathic hyperostosis, infectious and non-infectious inflammatory or destructive tissue changes of the lumbar spine, presence of significant infectious disease or other severe disabling health problems, substance abuse, ongoing treatment for back pain by other health care providers, pregnant or breast-feeding women, a pain score of less than three points on an 11-box scale (0-10)[27] or subject to pending or current litigation.

These criteria were implemented in order to achieve a homogeneous and stable cohort with consistent symptoms and severity of LBP, where the level of pain intensity made it possible to measure changes over time and where other health conditions were unlikely to influence the present outcomes during the one year clinical follow-up period.

Randomization and blinding

In the original study, restricted randomization using a 1:1:1 allocation ratio was applied using four strata: patients with and without radiating symptoms, LBP of 6 to 12 weeks' duration, and LBP for more than 12 weeks. Prior to enrollment, the project statistician generated a randomization list using randomly mixed permuted blocks of different sizes. Objective outcome assessment was performed by examiners masked to treatment allocation. Detailed information on randomization, recruitment and blinding procedures is reported in another publication [22].

Interventions (12 weeks)

Clinicians used standardized forms to document the events and procedures of each treatment visit, including patient-reported side-effects. A minimum of 80% attendance at their scheduled visits was required. The following intervention modalities were employed:

Spinal manipulative treatment

The number of treatments and the schedule of care were determined by one of the nine treating chiropractors. Treatment typically involved two encounters per week lasting 15-30 minutes that could include manual spinal manipulation, with light soft tissue massage, with the assistance of a flexion/distraction table if required. Activity modification was prescribed as necessary. The vertebral levels treated were determined by the individual clinicians by static and/or motion palpation. Specific spinal manipulation was performed as follows: patients were positioned on a treatment table in either the prone, supine or side-lying position. For each spinal manipulation, the chiropractor's contact hand would be placed over an osseous process, muscle or ligament and the vertebral or sacroiliac joint of interest would be taken to the end of its physiological ROM. The chiropractor would then apply a high velocity, low amplitude impulse to the joint, carrying it beyond the normal physiological ROM. Participants were discharged from care if the treating

clinician felt that maximum clinical benefit had been obtained i.e. the clinician judged there to be insufficient biomechanical dysfunction to continue manipulative treatments.

Supervised exercise therapy

Supervised high dose exercise in small groups of patients (3 to 4) was provided (one-on-one supervision) by 15 exercise therapists trained in the study protocol. The main focus was dynamic trunk strengthening exercises (trunk extensions and leg extensions) and abdominal exercises using low-tech methods. In addition, a core strengthening program and static stretches (series of six) with a focus on the lumbar, gluteal, and hamstring musculature before and after strengthening. Each stretch was done once, with the patients instructed to hold each stretch for three deep breaths. Over the 12-week period, patients were asked to attend 20 one-hour sessions involving a high number of repetitions (two to three sets of 15–30 repetitions for each exercise) and a progressive increase in muscle load (achieved by altering the patient's center of gravity when possible). The patients were instructed to perform repetitions until they could no longer do so using proper form. The study protocol was based on the dynamic trunk strengthening protocol described by Manniche [28] which includes trunk extensions and leg extensions and has, in part, been tested in a previous trial [29].

Home exercise and advice

Eleven therapists trained in the study protocol provided counseling on self-care education. Two one-hour sessions were conducted on self-care measures and ergonomics associated with work and activities of daily living. These included postural instructions and practical demonstrations of proper body mechanics performed with patient participation.

A more comprehensive description of the various intervention modalities has been published elsewhere [22].

Outcomes and measurements

A comprehensive description and analysis has been undertaken of patient-rated outcomes as well as descriptions of instrumentation, attachment, measurement procedure, data processing and analysis, and reliability [16;22].

In short, evaluation was conducted during baseline assessment and 4, 12 weeks after randomization. Self-report questionnaires were completed at each time point, independent of study providers and investigators. Objective outcome assessments (lumbar kinematics including both sagittal and coronal plane motions, as well as rotation and circumduction, trunk muscle strength, and endurance) were collected by a blinded examiner at both baseline assessments and at Week 12. As described in detail previously [16] regional lumbar motion were sampled during a standardized motion test using a six-degrees-of-freedom instrumented spatial linkage system with a sampling rate of 100 Hz (CA 6000 Spine Motion Analyzer; OSI, Union City, CA, USA) (Figure 1). In a recent literature review it was concluded that most instruments (including the CA 6000) used under standardized conditions may be considered reliable enough to be used for research purposes on the group level [30].

Fig. 1 Illustrating the CA 6000 Spine Motion Analyzer attached to a person in neutral position

The instrument was calibrated against an inclinometer at the beginning of each test day and zero-setting was performed for each subject in the neutral position before their first test. Each participant wore a loose T-shirt and trousers. The instrument was attached to the patient when standing in a neutral position with arms relaxed. The fixed extremity of the device was mounted on the sacral crest (S2) using a manufacturer-supplied belt. The mobile end of the device was mounted at the level of T7 using a manufacturer-supplied chest harness and the top edge of the horizontal metal

pieces was aligned evenly with the inferior angles of the scapulae (which is level with the T7 spinous process). The pelvic harness was applied so that the binding posts were level with the posterior superior iliac spines. Neutral position was defined as the patient standing with eyes open, facing forward, with the feet positioned a shoulder width apart and arms hanging freely at their side with the low back in a comfortable position that they can return to. For all test directions, stringent test instructions were verbally explained to the patients.

For backward and forward bending each patient received the following verbal explanation . *“Ok, I’ll have you find a neutral position for your low back. Place your arms across your chest and bend backwards from the waist as far as you can go. As you return to neutral, move your palms to your thighs and while sliding your palms down your legs, bend forward from the waist as far as you can go, and then return to neutral (arms across chest). It should be done at your own pace and without pausing”*.

For circumduction motion (full turning of the back), each patient received the following verbal explanation. *“Find your neutral position and look forward with your hands on your hips. First bend backwards, then roll to your left, forwards, to your right, to the back and return to neutral. It is important to go as far as you can go in all directions. This entire movement should be done at your own pace without pausing. So, it should look like this”*. After these trials the patients were asked to circumduct their back in the opposite direction and each patient received the following verbal explanation. *“Ok, this is the last one. It is the same as the one you just did, but you’ll go in the opposite direction. So go backwards, right, forwards, left, back and then back to neutral”*.

Each patient then performed several trial runs as a ‘warm up’. Two recordings were obtained at each test session that needed to display a total ROM variability of four degrees or less.

A custom-made MatLab program was used to reduce the 3D data into single numbered motion parameters [16].

The following motion parameters were determined:

1. *ROM (degree)* was calculated as the total angular range of regional lumbar motion in the sagittal plane expressed in degrees from maximum extension to maximum flexion ((Intraclass correlation coefficient ($ICC_{(1,1)}$)[31]) = 0.69) (Figure 2)
2. *Maximum flexion velocity (degree/second)* was calculated as the peak angular speed in the forward bending motion reached from full extension position to full flexion position ($ICC_{(1,1)} = 0.69$).
3. *Phase-plot area (degree²/second)* was defined as the area comprised by the phase-plot of sagittal flexion-extension angular motion versus velocity. Phase-plot area was calculated based on cross-product calculations between vectors drawn from the neutral position to each coordinate point in the phase-plot [16] ($ICC_{(1,1)} = 0.74$).
4. *Jerk Index* was calculated from maximum extension to maximum flexion position as the mean spectral frequency of the first derivative of the angular acceleration signal multiplied by movement duration [16]. This parameter indicates the number of changes in acceleration, i.e. the smoothness of the motion ($ICC_{(1,1)} = 0.55$).
5. *3D circumduction area (cm²)* was defined as the curved 3D surface formed by the translatory motion from the point (0,0,0) to each point formed by (x, y, z) coordinates. The area was calculated based on cross-product calculations between vectors drawn from the neutral position to each coordinate measurement point in the circumduction motion. The average of left and right circumduction areas was used in the analysis. [16] ($ICC_{(1,1)} = 0.68$).
6. *2D circumduction area (degree²)* was defined as the 2D surface area of the angular phase-plot formed by the frontal and sagittal motion (Figure 2). The area was calculated based on cross-product calculations between vectors drawn from the neutral position to each

coordinate measurement point in the circumduction motion. The average of left and right circumduction areas was used in the analysis [16] ($ICC_{(1,1)} = 0.81$).

The $ICC_{(1,1)}$ values presented after each measurement represent the reliability of regional lumbar motion parameter in pain intensity-stable participants ($n=149$) with a pain level defined as a maximum change of ± 1 on ordinal 11-box scale during the previous week between the baselines 7 to 14 day apart. As a guideline, Portney and Watkins suggest that values below 0.50 represent poor reliability, coefficients from 0.50 to 0.75 suggest moderate reliability, and values above 0.75 are indicative of good or high reliability [32]. This instrument has been evaluated in several studies for reliability [30] and validity [33]. In a recent literature review it was concluded that these instruments used under standardized conditions may be considered reliable enough to be used for research purposes on the group level [30].

Fig. 2

Clockwise circumduction motion in a typical patient before and after treatment. The area increased after treatment.

Statistical analysis

We first determined if those study participants who completed the motion tests were different from those who did not. Independent t-tests or chi-square tests were conducted on the following parameters: age, sex, BMI, duration of pain, baseline physical component score, baseline mental component score, baseline depression score (CESD), diagnostic group, LBP intensity and leg pain intensity (ordinal 11-box scale) [27], RMDQ [34] and intervention group.

All lumbar motion parameters except for ROM were non-normally distributed [35]. Various transformations were applied to the parameters and non-normal distribution still existed. Paired t-

test (ROM) or Wilcoxon signed rank test were used for comparison between paired data, and Wilcoxon Rank-Sum (Mann-Whitney U) test was used for comparison of unpaired data. For comparison of differences in pre-to-post changes between the three treatment groups the Kruskal-Wallis test was used.

The reliability and measurement error of the sagittal motion parameters has been reported elsewhere [16]. Reliability analysis for the circumduction motion parameters was calculated including ICC_(1,1) [31] and limits of agreement[36]. These analyses were based on logarithmic transformed data in order to fit formal statistical assumptions.

Results

A total of 630 individuals were evaluated for the study, of which 329 were excluded because they did not meet the exclusion criteria specified in the primary paper [22]. Therefore, 301 patients were recruited, but due to technical problems with the equipment at baseline or follow-up (80 patients) and dropouts (22 patients), a total of 199 complete patient recordings were obtained. Overall, adherence to study interventions was high i.e. the number of patients who did not receive or discontinued intervention for each treatment group were: 4 for HEA (refused to participate n=3, time commitment n=1), 4 for SMT (refused to participate n=3, competing co-morbidity n=1) and 14 for the SET (Unknown reason n=2, increase in pain n=3, refused to participate n=3, moved n=1, time commitment n=1, personal conflict n=3). More detail about the adherence is reported in the primary paper [22]. The individuals not available for analysis were significantly younger (Table 1) but there were no differences in other baseline characteristics such as BMI, gender, duration of pain, back/leg pain intensity and RMDQ score or depression score. Table 1 summarizes the demographic and clinical characteristics of the study participants. For the regional lumbar motion evaluation, 199 persons had complete motion data at baseline and Week 12. Of these, 62 received SET, 77 received SMT and 60 received HEA.

We hypothesized that the groups receiving either SET or SMT care would change significantly in all motion parameters over a 12-week period, whereas the minimal intervention group (HEA) would not (no change in motion parameters). Specifically the Jerk index was hypothesized to decrease and all other motion parameters to increase. The SMT group increased on all parameters except for the Jerk Index, which decreased significantly. The two exercise groups increased on 3 out of 6 motion parameters (Table 4). The pre-to-post change in Jerk Index differed between treatments ($p = 0.0031$), with the SMT group changing to a smoother motion. Therefore we could not confirm our hypothesis.

Discussion

Using a 3D regional lumbar motion instrument, we tested the theory that in chronic LBP patients, commonly used treatments actually change regional lumbar motion by modulating spinal biomechanics, and that specific treatment modalities may affect regional lumbar motion differently. The results presented in the current study suggest that changes in regional lumbar motion can occur in chronic LBP patients following high dose (SET) well as low dose (HEA) exercise and spinal manipulation, and these changes are potentially different between the interventions.

The regional lumbar motion parameters examined in the current study were chosen in order to obtain a more complete representation of regional lumbar motion biomechanics than achieved by sagittal-plane ROM alone. These motion parameters can be divided into time-independent and time-dependent parameters. The time-independent parameters examined in this study were sagittal ROM and two circumduction motion areas. Circumduction areas were calculated using all 2D or 3D data points measured during the circumduction motion and may therefore represent a more relevant measure than single-plane ROM when quantifying functional impairments. In a clinical setting, such circumduction measurements may be useful and practical because one simple test gives an impression of a patient's overall functional ability by examining movement in several directions in

the same manoeuvre. Besides the numerical calculation of the motion area, visual tools such as depicted in Figure 2 may prove useful in the clinical setting. Such information, including the visual tools, might assist diagnosis and the evaluation of changes between successive treatment visits. Although the present 2D circumduction motion parameters encompass a complex motion scenario, the reliability seems to be better than for other parameters used in this study. Thus, with relevant protocol adjustments and/or technological improvements, it may be useful at an individual patient level in clinical settings.

Research has indicated that LBP [15] or neck pain patients [37] show reduced velocity and more jerky movements and that these variables therefore could change in predictable ways if the treatment is effective. Time-dependent parameters (phase-plot area, velocity and the Jerk Index) are less static psychometric measures expressing dynamic regional lumbar motion characteristics using higher order derivatives. In order to condense this information into single metric values containing as much kinematic data as possible, the phase-plot area was calculated using the combined sagittal angular position and angular velocity signals, respectively.

The hypothesis that LBP patients would move more smoothly following SET or SMT treatment was not confirmed i.e. only the SMT group changed to a smoother motion. It has been hypothesized that unconscious motor control mechanisms rather than psychological processes might be responsible for the increased Jerk Index in chronic neck pain patients, although this remains to be clarified [37]. The same mechanisms might be present in chronic LBP patients. Assuming that the Jerk Index, to some extent, reflects neuromuscular motor control strategies in chronic LBP patients, it is quite interesting that SMT and SET seem to have opposite effects on these motor control mechanisms (Figure 3). These findings may therefore support the theory that these treatments partly work through different biological pathways e.g. neuromuscular versus hormones, such as endorphins. Following this line of thinking, these treatments could support each other in terms of

optimizing patient outcomes e.g. as hypo-analgesic treatment effect and physical functioning. The Jerk Index is limited in that it had relatively low reliability i.e. much noise relative to signal. However this makes it even more surprising that a significant difference ($p=0.0031$) between treatment groups was found. This index can be calculated in different ways and other formulae may provide more reliable outcomes, which should be explored in future studies.

To facilitate their use in the clinical setting, these parameters need to be studied more closely to determine their sensitivity to assist diagnosis or to assess therapeutic patient outcomes.

Study Limitations / Strengths

This study examined a large number of participants, with a relatively stable LBP level at baseline and a high degree of chronicity (Table 1). In general, the objectivity of all regional lumbar motion measurements can be questioned, for example, a patient may exaggerate movements for a variety of purposes, either consciously or unconsciously. However, in the current study, patients with ongoing pending or current litigation were excluded. In addition, the learning effect is a well-known phenomenon that may influence an outcome i.e. change the course of movements in the habituated state. To minimize this potential problem, all patients participated in two baseline assessments and all analyses were done using data from the second baseline only.

The fixed extremity of the CA6000 Spine Motion Analyzer was mounted on the sacral crest (S2) and the mobile end of the device was mounted at the level of T7 spinous process. Therefore the measured movements were due both lumbar motion and motion from the lower part of the thoracic region.

Since these data were from a randomized comparative study, we believe that the current conclusions about different treatments are fairly robust. However, the absence of a strict no-treatment control group raises the possibility that the changes currently observed reflect the natural change in regional lumbar motion characteristics over time.

Initially, 301 patients were recruited but due to technical problems and drop outs, only 199 complete patient recordings were obtained at follow-up. The individuals not available for analysis were slightly younger but there were no differences in other baseline characteristics such as BMI, gender, duration of pain, or depression score.

The complexity of spinal motion is enormous and in order to use recordings as a quantitative outcome, data had to be condensed into a manageable number of parameters, which may have resulted in potentially important information being lost. In addition, we did not assess short-term regional lumbar movement changes or immediate and short-term treatment effects. Finally, the study did not address whether the observed changes in regional lumbar motion outcomes were translated to improved patient outcomes, nor if particular patterns of baseline motion characteristics were able to predict the range of adaptive improvement. These important aspects should be addressed in future studies.

Conclusions

This study provides evidence that regional lumbar motion changes can occur in chronic low back pain patients over a 12-week intervention period. Treatments in the form of exercise or spinal manipulation appear to elicit dissimilar adaptive changes in regional lumbar motion ability.

Reference List

- [1] Andersson GBJ. Epidemiological features of chronic low-back pain. *Lancet* 1999;354:581-5.
- [2] Manek NJ, MacGregor AJ. Epidemiology of back disorders: prevalence, risk factors, and prognosis. *Curr Opin Rheumatol* 2005;17:134-40.
- [3] Maus T. Imaging the back pain patient. *Phys Med Rehabil Clin N Am* 2010;21:725-66.
- [4] Maus TP. Imaging of the spine and nerve roots. *Phys Med Rehabil Clin N Am* 2002;13:487-544, vi.
- [5] Beattie PF, Meyers SP. Magnetic resonance imaging in low back pain: general principles and clinical issues. *Phys Ther* 1998;78:738-53.
- [6] Jarvik JG, Deyo RA. Diagnostic evaluation of low back pain with emphasis on imaging. *Ann Intern Med* 2002;137:586-97.
- [7] Kjaer P, Leboeuf-Yde C, Korsholm L, Sorensen JS, Bendix T. Magnetic resonance imaging and low back pain in adults: A diagnostic imaging study of 40-year-old men and women. *Spine* 2005;30:1173-80.
- [8] Deyo RA. Diagnostic evaluation of LBP - Reaching a specific diagnosis is often impossible. *Arch Intern Med* 2002;162:1444-7.
- [9] Lehman GJ. Biomechanical assessments of lumbar spinal function. How low back pain sufferers differ from normals. Implications for outcome measures research. Part I: kinematic assessments of lumbar function. *J Manipulative Physiol Ther* 2004;27:57-62.
- [10] Marras WS, Parnianpour M, Ferguson SA et al. The Classification of Anatomic-Based and Symptom-Based Low-Back Disorders Using Motion Measure Models. *Spine* 1995;20:2531-46.
- [11] Tousignant M, Poulin L, Marchand S, Viau A, Place C. The Modified-Modified Schober Test for range of motion assessment of lumbar flexion in patients with low back pain: A study of criterion validity, intra- and inter-rater reliability and minimum metrically detectable change. *Disabil Rehabil* 2005;27:553-9.
- [12] Nattrass CL, Nitschke JE, Disler PB, Chou MJ, Ooi KT. Lumbar spine range of motion as a measure of physical and functional impairment: an investigation of validity. *Clin Rehabil* 1999;13:211-8.
- [13] Sullivan MS, Shoaf LD, Riddle DL. The relationship of lumbar flexion to disability in patients with low back pain. *Phys Ther* 2000;80:240-50.

- [14] Panjabi MM, Oxland TR, Yamamoto I, Crisco JJ. Mechanical-Behavior of the Human Lumbar and Lumbosacral Spine As Shown by 3-Dimensional Load-Displacement Curves. *J Bone Joint Surg Am* 1994;76A:413-24.
- [15] McGregor AH, McCarthy ID, Hughes SPF. Motion Characteristics of Normal Subjects and People with Low Back Pain. *Physiotherapy* 1995;81:632-7.
- [16] Mieritz RM, Bronfort G, Jakobsen MD, Aagaard P, Hartvigsen J. Reliability and measurement error of regional lumbar motion parameters in 220 chronic low back pain patients. *Spine J* 2012;submitted.
- [17] McGregor AH, Cattermole HR, Hughes SP. Spinal motion in lumbar degenerative disc disease. *J Bone Joint Surg Br* 1998;80:1009-13.
- [18] McGregor AH, Hughes SPF. The potential use of spinal motion as a measure of surgical outcome. *J Back Musculoskelet* 2004;17:77-82.
- [19] Evans DW. Mechanisms and effects of spinal high-velocity, low-amplitude thrust manipulation: Previous theories. *J Manipulative Physiol Ther* 2002;25:251-62.
- [20] Shekelle PG. Spine Update - Spinal Manipulation. *Spine* 1994;19:858-61.
- [21] Bronfort G, Haas M, Evans R, Kawchuk G, Dagenais S. Evidence-informed management of chronic low back pain with spinal manipulation and mobilization. *Spine J* 2008;8:213-25.
- [22] Bronfort G, Maiers MJ, Evans RL et al. Supervised exercise, spinal manipulation, and home exercise for chronic low back pain: a randomized clinical trial. *Spine J* 2011;11:585-98.
- [23] Hayden JA, van Tulder MW, Tomlinson G. Systematic review: Strategies for using exercise therapy to improve outcomes in chronic low back pain. *Ann Intern Med* 2005;142:776-85.
- [24] Hayden JA, van Tulder MW, Malmivaara AV, Koes BW. Meta-analysis: Exercise therapy for nonspecific low back pain. *Ann Intern Med* 2005;142:765-75.
- [25] Riddle DL. Relationships between disability and physical impairment in patients with low back pain. Virginia Commonwealth University Richmond, Virginia; 1997.
- [26] Atlas SJ, Deyo RA, Patrick DL, Convery K, Keller RB, Singer DE. The Quebec Task Force classification for spinal disorders and the severity, treatment, and outcomes of sciatica and lumbar spinal stenosis. *Spine* 1996;21:2885-92.
- [27] Jensen MP, Karoly P, Braver S. The measurement of clinical pain intensity: a comparison of six methods. *Pain* 1986;27:117-26.
- [28] Manniche C, Bentzen L, Hesselsoe G, Christensen I, Lundberg E. Clinical-Trial of Intensive Muscle Training for Chronic Low-Back Pain. *Lancet* 1988;2:1473-7.
- [29] Bronfort G, Goldsmith CH, Nelson CF, Boline PD, Anderson AV. Trunk exercise combined with spinal manipulative or NSAID therapy for chronic low back pain: a randomized, observer-blinded clinical trial. *J Manipulative Physiol Ther* 1996;19:570-82.

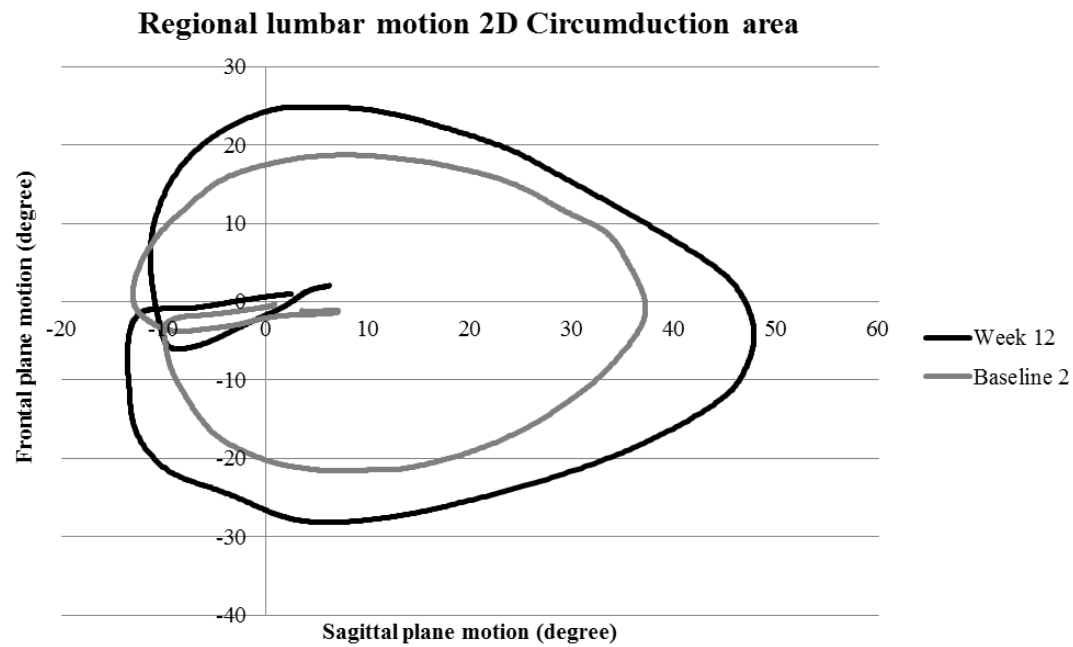
- [30] Mieritz RM, Bronfort G, Kawchuk G, Breen A., Hartvigsen J. Reliability and measurement error of 3D regional lumbar motion measures: a systematic review. *J Manipulative Physiol Ther* 2012;35:645-56.
- [31] Shrout PE, Fleiss JL. Intraclass Correlations - Uses in Assessing Rater Reliability. *Psychol Bull* 1979;86:420-8.
- [32] Portney LG, Watkins M.P. *Foundation of clinical research: applications to practice*. Third ed. Upper Saddle River, New Jersey: Pearson Education, Inc.; 2009.
- [33] Mannion A, Troke M. A comparison of two motion analysis devices used in the measurement of lumbar spinal mobility. *Clin Biomech* 1999;14:612-9.
- [34] Patrick DL, Deyo RA, Atlas SJ, Singer DE, Chapin A, Keller RB. Assessing Health-Related Quality-Of-Life in Patients with Sciatica. *Spine* 1995;20:1899-908.
- [35] Friedman LM, Furberg CD, DeMets DL. *Fundamentals of clinical trials*. 4th Edition ed. New York: Springer; 2010.
- [36] Bland JM, Altman DG. Measuring agreement in method comparison studies. *Stat Methods in Med Res* 1999;8:135-60.
- [37] Sjolander P, Michaelson P, Jaric S, Djupsjobacka M. Sensorimotor disturbances in chronic neck pain - Range of motion, peak velocity, smoothness of movement, and repositioning acuity. *Man Ther* 2008;13:122-31.

Figure 1



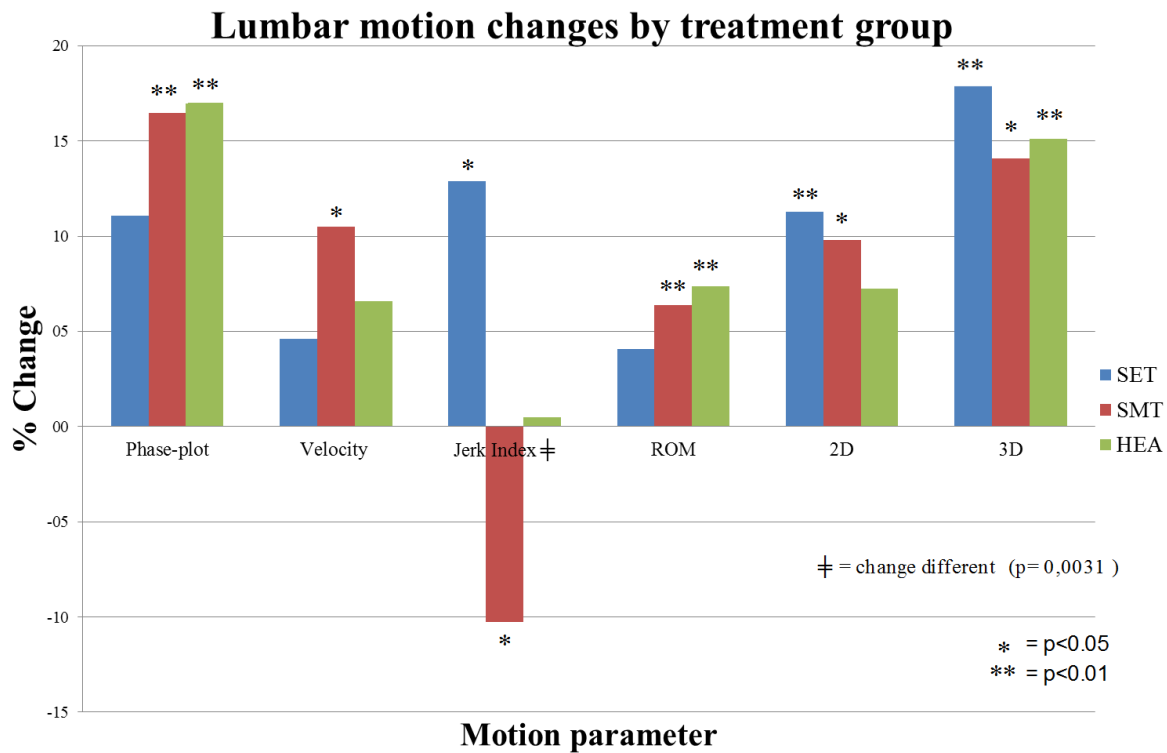
Illustrating the CA6000 Spine Motion Analyzer attached to a person in neutral position

Figure 2



Clockwise circumduction motion in a typical patient before and after treatment. The area increased after treatment.

Figure 3



Lumbar motion percentage changes between baseline and 12-week follow up by motion parameters and treatment groups. SET = Supervised exercise therapy, SMT = Spinal manipulative therapy, HEA = Home exercise and advice, Phase-plot = Phase-plot Area (degree²/sec), Velocity = Maximum flexion velocity (degree /sec), Jerk Index = number of changes in acceleration from full extension to full flexion, ROM= Range of motion (degree), 2D = 2-dimensional circumduction area (degree²), 3D = 3-dimensional circumduction area (cm²). ROM p values using paired t-test. All others using Wilcoxon signed rank test. † change different for SMT vs. SET and HEA (Kruskal-Wallis test). NB: This figure is based on the data and statistics presented in Table 2. It is made for descriptive purposes and interpretations should be done carefully.

Figure 4



The CA6000 Spine Motion Analyzer with a person in neutral, extension and flexion positions

Figure 5



Illustration of a sequence of pictures (read from top left to bottom right) of the CA6000 Spine Motion Analyzer attached to a person performing circumduction motion (neutral, extension, left lateral flexion, flexion, right lateral flexion, extension and back to neutral position).

Table 1.

Baseline characteristics for 199 chronic LBP patients participating in a randomized clinical trial comparing the relative effectiveness of spinal manipulative treatment, supervised exercise therapy, and home exercise.

Intervention group (n)	SMT (77)		SET (62)		HEA (60)		ALL (199)		NMD (80)	
Characteristic (n=199)	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Age (years)	46	11	45	12	47	11	46	11	44*	10
BMI	28	6	28	6	28	5	28	6	28	5
Males (%)	35		45		40		40		40	
<u>Pain level (0-10)</u>										
Baseline 1	5.3	1.5	5.0	1.4	5.1	1.6	5.1	1.5	5.5	1.7
Baseline 2	5.3	1.8	5.0	1.5	5.1	1.8	5.2	1.7	5.2	1.7
Pain more than 1 year (%)	84		77		83		82		83	
<u>Quebec Task Force Classification (%)</u>										
1	69		66		67		67		67	
2	19		18		17		18		18	
3	12		15		12		13		10	
4			2		5		2		5	
SF36 Physical Component	43	7	44	7	43	8	43	7	42	8
SF36 Mental Component	55	8	54	8	54	8	54	8	54	8
Depression (CESD)	10	10	13	11	11	9	11	10	13	10
Roland Morris	36	20	36	21	37	24	36	21	37	21

* Individuals who did not complete were significantly different

§ Depression is defined as greater than 16 points on the Center for Epidemiologic Studies Depression Scale (CESD)

(SMT) = Spinal manipulative treatment, (SET) = Supervised exercise therapy and (HEA) = Home exercise and advice

(NMD) = no motion data due to instrument error

Table 2.

Lumbar motion characteristics at baseline and 12-week follow up by treatment group

Supervised exercise therapy (SET)									
(n=62)		Baseline			Follow Up 12 Weeks			p value	
Spinal Measure	Mean	95% CI			Mean	95% CI			
<u>Flexion</u>									
Phase-plot	4038	3611	-	4465	4485	3956	-	5013	0.102
Velocity	36.8	33.8	-	39.7	38.5	34.6	-	42.5	0.477
Jerk Index	10.1	9.2	-	11.1	11.4	10.1	-	12.6	0.021
ROM	71.0	67.2	-	74.8	73.9	69.9	-	78.0	0.092
<u>Circumduction</u>									
2D	2722	2448	-	2996	3029	2704	-	3353	0.010
3D	144	125	-	163	170	145	-	195	0.004

Spinal manipulative therapy (SMT)									
(n=77)		Baseline			Follow Up 12 Weeks			p value	
Spinal Measure	Mean	95% CI			Mean	95% CI			
<u>Flexion</u>									
Phase-plot	4018	3601	-	4434	4680	4175	-	5185	0.003
Velocity	36.2	32.8	-	39.6	40.0	36.3	-	43.8	0.046
Jerk Index ‡	10.7	9.5	-	11.9	9.6	8.7	-	10.6	0.014
ROM	69.0	65.5	-	72.5	73.4	69.4	-	77.4	0.011
<u>Circumduction</u>									
2D	2714	2442	-	2985	2980	2653	-	3307	0.053
3D	135	117	-	153	154	134	-	174	0.028

Home exercise and advice (HEA)									
(n=60)		Baseline			Follow Up 12 Weeks			p value	
Spinal Measure	Mean	95% CI			Mean	95% CI			
<u>Flexion</u>									
Phase-plot	3884	3396	-	4372	4544	3966	-	5123	0.005
Velocity	34.9	31.6	-	38.2	37.2	33.9	-	40.6	0.169
Jerk Index	9.9	8.8	-	10.9	9.9	8.9	-	11.0	0.988
ROM	68.0	63.0	-	73.0	73.0	68.0	-	78.1	0.005
<u>Circumduction</u>									
2D	2727	2384	-	3070	2924	2540	-	3308	0.069
3D	139	116	-	162	160	134	-	187	0.006

Phase-plot = Phase-plot Area (degree²/sec), Velocity = Maximum flexion velocity (degree/sec), Jerk Index = number of changes in acceleration from full extension to full flexion, ROM= Range of motion (degree), 2D = 2-dimensional circumduction area (degree²), 3D = 3-dimensional circumduction area (cm²). ROM p values using paired t-test. All others using Wilcoxon signed rank test. ‡ change different for SMT vs. SET and HEA (Kruskal-Wallis test p= 0,0031).

Manuscript IV

Regional lumbar motion and patient-rated outcomes

A secondary analysis of data from a randomized clinical trial

Rune M. Mieritz, DC, M.Sc.¹, Gert Bronfort, DC, PhD^{2,3}, Jan Hartvigsen, DC, PhD^{1,2}

¹Institute of Sports Science and Clinical Biomechanics, University of Southern Denmark, Odense, Denmark; ²Nordic Institute of Chiropractic and Clinical Biomechanics, Odense, Denmark;

³Northwestern Health Sciences University, Minneapolis, USA

Corresponding Author:

Rune Mygind Mieritz

Institute of Sports Science & Clinical Biomechanics, University of Southern Denmark

Campusvej 55, DK-5230 Odense M, Denmark

PH: 0045 61 308 589

Mail: rmieritz@health.sdu.dk

Word count abstract: 301

Word count main text: 4714

Number of figures: 1

Number of tables: 6

Keywords: low back pain, spine, lumbar, measurement, assessment, motion analysis, biomechanics, range of motion, nonsurgical, treatment, manipulation, exercise.

Abstract

Background: Health professionals routinely measure lumbar motion in patients presenting with low back pain (LBP) and such measurements may be used for monitoring the biomechanical status of the low back, choosing a treatment modality and evaluating the treatment's effectiveness. The purpose of the current study was to examine the relationship in change scores between regional lumbar motion and patient-rated pain of the previous week and back-related function in chronic LBP patients enrolled in a randomized clinical trial and treated with either exercise therapy or spinal manipulation using six different motion parameters.

Methods: Regional lumbar motions were sampled by trained therapists using a six-degrees-of-freedom instrumented spatial linkage system in 199 participants at baseline and 12 weeks follow up. The regional lumbar motion data were analyzed as a total cohort as well as relative to subgroup stratifications; back pain only versus back and leg pain, and treatment modality (supervised exercise, spinal manipulation and home exercises). For identifying clinically meaningful improvements in the measurements of back pain and back-related function, we used a 30% threshold.

Results: The relationship between change scores in patient-rated outcomes and objective measures of regional lumbar motion was found to be weak (ranging from no correlation to -0.37). In contrast, distribution of pain and treatment received affected associations between motion parameters and patient-rated outcomes. Thus stronger correlation coefficients and significant differences between clinically relevant improved versus no clinical relevant change were found in some motion parameters in the subgroup with back pain only and the treatment group receiving spinal manipulation.

Conclusion: Overall changes in regional lumbar motion were poorly associated with patient-rated outcomes measured by back-related function and back pain intensity scores. However, associations between regional lumbar motion versus patient-rated pain and back-related function were different relative to subgroups i.e. back pain only versus back and leg pain and treatment.

(301 words)

Introduction

Health professionals routinely measure regional lumbar motion in patients presenting with low back pain (LBP) in search of objective findings that could potentially explain the patient's pain. Such findings may be used for assessing the severity of the condition, guiding treatment decisions and monitoring treatment effectiveness¹⁻⁴. The theory behind these assumptions is that impairment in the form of restriction in regional lumbar motion is a contributor to loss of back-related function or pain reported by the patient. Several studies have used objective measures such as range of motion (ROM) to test the nature of these theoretical relationships⁵⁻¹³.

Commonly used clinical measures for patient-rated outcomes are pain scales and questionnaire-based instruments assessing back-related function such as the Roland Morris Disability Questionnaire (RMDQ). More objective functional assessments of lumbar motion can be achieved by a variety of methods including examination of regional motion using simple low-tech measurements such as inclinometer ROM, finger-to-floor distance or Schober's Index. However associations between these low-tech ROM measures and the patient-rated measures are low^{9;14-16}. However, the development of new technologies has made it possible to derive and quantify more sophisticated motion parameters such as motion velocity, acceleration, symmetry or motion area in 3D^{5;11;17;18}. These new motion parameters show more promise for quantifying a patient's functional level and may correlate well with patient-rated measures of pain and disability, in particular when used in a longitudinal context even though there is some variation between studies^{6;7;19;20}.

The overall aim of the current study was to examine, using six different motion parameters, the relationship between regional lumbar (from S1 to T7 spinous process) motion and patient-rated outcomes in a sample of 199 chronic LBP patients enrolled in a randomized clinical trial (RCT).

Specifically, we wanted to:

- 1) Report associations and compare change between six different motion parameters (sagittal spinal ROM, maximum flexion velocity, phase-plot area, Jerk Index (smoothness of motion) and two circumduction area motion parameters) with change in back pain level and change in back-related function (RMDQ) over a 12-week period, for the whole study sample and for a sub-sample of participants, excluding those with the highest baseline motion scores, as there was less potential for improvement.

We hypothesized that excluding patients with the highest baseline scores would result in stronger associations (hypothesis 1). We also hypothesized that if patients had a clinically relevant improvement in pain and back-related function (i.e. >30%), this would correspond to greater change in the motion parameters compared to patients who did not achieve a clinically relevant improvement (hypothesis 2).

- 2) Exploratorily investigate whether this relationship was similar in subgroups based on: (i) pain distribution, (back pain only versus back and leg pain), and (ii) receiving different treatments (spinal manipulation, supervised trunk exercise, or home exercises over a 12-week period).

Material and Methods

Design

The research design was a prospective cohort study of participants from an observer-blinded, parallel-group, RCT²¹. The participants were consecutively recruited over a period of 3 years at the Wolfe Harris Center for Clinical Studies at Northwestern Health Sciences University, Minneapolis, USA. The institutional review boards of the Northwestern Health Sciences University, the Minneapolis Medical

Research Foundation, and the University of Minnesota approved the study. Written informed consent was obtained from all participants. Regional lumbar motion recordings were measured during two baseline visits (separated by 7-14 days) and one follow-up visit after 12 weeks of intervention. To illustrate the stability of pain intensity in the overall cohort, pain intensity levels recorded at the two baseline measurement time points are presented in Table 1.

Participants

Inclusion/exclusion criteria

The participants must have completed baseline and follow-up regional lumbar motion assessment procedures before being randomized in the RCT. Additional inclusion criteria were: 18-65 years of age, and a primary complaint of mechanical LBP of at least 6 weeks' duration with or without radiating pain to the lower extremity. Mechanical LBP was defined as pain that had no specific identifiable etiology but that could be reproduced by back movements or provocation tests. Exclusion criteria were: previous lumbar spine fusion surgery, progressive neurological deficits, aortic or peripheral vascular disease, pain scores of less than 3 (0–10 scale), ongoing treatment for back pain by other health care providers, or participation in pending or current litigation. Participants were recruited through local newspaper advertisements, community posters, and postcard mailings and initial screening was conducted by telephone.

Randomization and blinding

In the original study, restricted randomization using a 1:1:1 allocation ratio was applied using four strata: LBP with radiating symptoms, LBP without radiating symptoms, LBP of 6-12 weeks' duration, and LBP >12 weeks. Objective outcome assessment was performed by examiners masked to treatment allocation. Detailed information on randomization, recruitment and blinding procedures have been previously reported ²¹.

Interventions (12 weeks)

The participating clinicians used standardized forms to document the events and procedures of each treatment visit, including patient-rated side-effects. A minimum of 80% attendance at their scheduled visits was required. The following intervention modalities were employed:

Spinal manipulative treatment (SMT)

The number of treatments and the schedule of care were determined by each of the nine treating chiropractors. Treatment typically involved two encounters per week lasting 15-30 minutes that could include manual spinal manipulation with light soft tissue massage and table-assisted flexion/distraction and/or prescribed activity modification as necessary. The vertebral levels treated were determined by the individual clinicians using static and/or motion palpation. The specific spinal manipulation procedures used have been previously reported ²¹.

Supervised exercise therapy (SET)

Supervised high dose exercise in small groups of patients (3 to 4) was provided (one-on-one supervision) by 15 exercise therapists trained in the study protocol. The main focus was dynamic trunk strengthening exercises (trunk extensions and leg extensions) and abdominal exercises using low-tech methods. In addition, a core strengthening program and static stretches (series of six) with a focus on the lumbar, gluteal, and hamstring musculature before and after strengthening were implemented. More detail on the exercise therapy used has been previously reported ²¹.

Home exercise and advice (HEA)

Eleven therapists trained in the study protocol provided counseling on self-care education. The HEA group was intentionally minimal in its approach so it could serve as control ²¹. Two one-hour sessions were conducted on self-care measures and ergonomics associated with work and activities of daily living. These included postural instructions and practical demonstrations of

proper body mechanics performed with patient participation. A more comprehensive description of the various intervention modalities has been published elsewhere ²¹.

Outcomes and measurements

A comprehensive description and analysis has been previously reported of the patient-rated outcomes

²¹. Briefly, an evaluation was conducted during the baseline assessment and 12 weeks after randomization. Patient-rated questionnaires were completed at each time point, independent of study providers and investigators. Pain was measured on an ordinal 11-box scale ²² where the patients were asked to rate their typical level of back pain over the previous week on a 0-10 scale, with 0 being ‘no pain’ and 10 being ‘worst pain possible’ and Modified Roland Morris Disability Questionnaire (RMDQ) ²³ was completed at each time point, independent of study care providers and investigators.

Objective outcome assessments (regional lumbar motion including both sagittal and coronal plane motions, as well as rotation and circumduction, trunk muscle strength, and endurance) were collected by a blinded examiner at each time point. Regional lumbar motion data were sampled during a standardized motion test using a six-degrees-of-freedom instrumented spatial linkage system with a sampling rate of 100 Hz (CA 6000 Spine Motion Analyzer; OSI, Union City, CA, USA).

The instrument was calibrated against an inclinometer at the beginning of each test day and zero-setting was performed for each subject in the neutral position before the first test. Participants wore loose T-shirts and trousers. The instrument was attached to the patient when standing in a neutral position with arms relaxed. The fixed extremity of the device was mounted on the sacral crest (S2) using a manufacturer-supplied belt. The mobile end of the device was mounted at the level of T7 using a manufacturer-supplied chest harness and the top edge of the horizontal metal pieces was aligned evenly with the inferior angles of the scapulae (which is level with the T7 spinous process). The pelvic

harness was applied so that the binding posts were level with the posterior superior iliac spines. Neutral position was defined as the patient standing with eyes open, facing forward, with the feet positioned a shoulder width apart and arms hanging freely at their side with the low back in a comfortable position. For all test directions, stringent test instructions were verbally explained to the patients.

For backward and forward bending each patient received the following verbal explanation . *“Ok, I’ll have you find a neutral position for your low back. Place your arms across your chest and bend backwards from the waist as far as you can go. As you return to neutral, move your palms to your thighs and while sliding your palms down your legs, bend forward from the waist as far as you can go, and then return to neutral (arms across chest). It should be done at your own pace and without pausing”.*

For circumduction motion (full turning of the back), each patient received the following verbal explanation. *“Find your neutral position and look forward with your hands on your hips. First bend backwards, then roll to your left, forwards, to your right, to the back and return to neutral. It is important to go as far as you can go in all directions. This entire movement should be done at your own pace without pausing. So, it should look like this”.* After these trials the patients were asked to circumduct their back in the opposite direction and each patient received the following verbal explanation. *“Ok, this is the last one. It is the same as the one you just did, but you’ll go in the opposite direction. So go backwards, right, forwards, left, back and then back to neutral”.*

Each patient then performed several trial runs as a ‘warm up’. Two recordings were obtained at each test session that needed to display a total ROM variability of four degrees or less. A project-specific MatLab computer program was used to reduce the 3D data into single numbered motion parameters.

The following motion parameters were determined:

1. *ROM (degree)* was calculated as the total angular range of lumbar motion in the sagittal plane expressed in degrees from maximum extension to maximum flexion (Intraclass correlation coefficient ($ICC_{(1,1)}$)= 0.69).
2. *Maximum flexion velocity (degree/second)* was calculated as the peak angular speed in the forward bending motion reached from full extension to full flexion ($ICC_{(1,1)} = 0.69$).
3. *Phase-plot area (degree²/second)* was defined as the area comprised by the phase-plot of sagittal flexion-extension angular motion versus velocity. Phase-plot area was calculated based on cross-product calculations between vectors drawn from the neutral position to each coordinate point in the phase-plot ($ICC_{(1,1)} = 0.74$).
4. *Jerk Index* was calculated from maximum extension to maximum flexion as the mean spectral frequency of the first derivative of the angular acceleration signal multiplied by movement duration. This parameter indicates the number of changes in acceleration, i.e. the smoothness of the motion ($ICC_{(1,1)} = 0.55$).
5. *Two-dimensional (2D) circumduction area (degree²)* was defined as the 2D surface area of the angular phase-plot formed by the frontal and sagittal motion (Figure 1). The area was calculated based on cross-product calculations between vectors drawn from the neutral position to each coordinate measurement point in the circumduction motion. The average of left and right circumduction areas was used in the analysis ($ICC_{(1,1)} = 0.81$).
6. *3D circumduction area (cm²)* was defined as the curved 3D surface formed by the translatory motion from the point (0,0,0) to each point formed by (x, y, z) coordinates. The area was calculated based on cross-product calculations between vectors drawn from the neutral position

to each coordinate measurement point in the circumduction motion. The average of left and right circumduction areas was used in the analysis ($ICC_{(1,1)} = 0.68$).

The $ICC_{(1,1)}$ values presented after each measurement represent the reliability of regional lumbar motion parameter in pain intensity-stable participants (n=149) with a pain level defined as a maximum change of ± 1 on ordinal 11-box scale during the previous week between the baselines 7 to 14 day apart. As a guideline, Portney and Watkins suggest that values below 0.50 represent poor reliability, coefficients from 0.50 to 0.75 suggest moderate reliability, and values above 0.75 are indicative of good or high reliability ²⁴. This instrument has been evaluated in several studies for reliability ²⁵ and validity ²⁶. In a recent literature review it was concluded that these instruments used under standardized conditions may be considered reliable enough to be used for research purposes on the group level ²⁵.

Statistical analysis

We first determined if the participants who completed the motion tests were different compared to those who did not, using non-paired t-tests or chi-square tests on the following parameters: age, sex, BMI (kg/m^2), duration of pain, baseline physical component score (SF-36), baseline mental component score (SF-36), baseline depression score (CESD), Quebec Task Force classification, LBP intensity and leg pain intensity (ordinal 11-box scale) ²², RMDQ ²³ and intervention group.

No significant difference between the three treatment groups SET, SMT and HEA in terms of pain and other patient-rated outcomes, in short- and long-term were found in the primary analysis ²¹. Based on these results we found it acceptable to collapse these treatment groups in order to analyses associations in changes in pain and back related function versus regional lumbar motion in the total cohort. In order to examine how this potential relationship relates to other factors we also did an exploratory analysis in

order to investigate relationships of the data relative to subgroups i.e. based on pain distribution, (back pain only versus back and leg pain), and different treatments (SMT, SET and HEA).

The learning effect is a well-known phenomenon that may influence outcome i.e. change the course of movements in the habituated state. To minimize this potential problem, all patients participated in two baseline assessments and all analyses were based on change scores between the second baseline and after 12 weeks of intervention. For identifying clinically meaningful improvements in the measurements of back pain and back-related function (RMDQ), we used a 30% threshold as recommended by Ostelo et al.²⁷.

An increase in motion score was considered to be related to an improvement in patient-rated outcome with the exception of the Jerk Index, which was expected to be the opposite i.e. the clinical improvement would result in a smoother motion and, thus, a lower score.

In order to assess whether a ceiling effect was present for the six regional lumbar motion parameters, we stratified participants into quartiles for these measurements in order to exclude those above the 75th baseline percentile, as they were unlikely to improve, as they already had proper motion at baseline. Systematic differences between age groups, genders and obesity levels have previously been reported^{28;29}. Therefore, based on the assumption that a ceiling effect could be present in all subgroups, these factors were adjusted for, by normalizing each subgroup relative to the subgroup mean prior to the stratification into quartiles. Age was adjusted based on 10-year age groups and obesity was adjusted using the cut-off point greater or equal to 30 kg/m² as recommended by the WHO³⁰. Differences in LBP intensity between the quartiles were examined.

All regional lumbar motion parameters except for ROM were non-normally distributed and we therefore used nonparametric statistics where possible and logarithmic transformed data in order to fit

formal statistical assumptions for calculation of Intraclass correlation coefficient. In addition we presented mean values for ROM but median values of all other regional lumbar motion parameters with 95% confidence intervals because of the non-normal distribution and calculated Spearman correlation coefficients. Patients who became worse (deteriorated) is also presented in tables but the groups were very small (ranging from 1 to 8 participants).

Results

Three hundred and one patients were recruited and randomized, but due to technical problems with the equipment at baseline and follow-up (80 patients) and also dropouts (22 patients), a total of 199 complete patient recordings were obtained at baseline and at week 12. Of these, 62 received supervised exercise therapy, 77 received spinal manipulative therapy and 60 received home exercises and advice. The individuals not available for analysis were significantly younger (based on confidence intervals) but there were no significant differences in the other baseline characteristics of BMI (kg/m^2), gender, duration of pain, depression score, back/leg pain distribution, back/leg pain intensity and RMDQ score (Table 1).

The change scores for the motion parameters from the baseline upper quartile group were radically different from change scores for the other quartiles in all motion parameters indicating a possible ceiling effect (4 out of 6 motion parameters were significant different from the other quartiles) (Table 2), however there were no significant differences in LBP intensity between the quartiles. Exclusion of patients from the baseline upper quartile group did not improve the correlation between motion parameter change scores and change scores for pain or back-related function as hypothesized (Table 3 A and B). Thus, our hypothesis that excluding the patients with the highest baseline scores would result in stronger correlations between change scores could not be confirmed (Table 3).

In general, low and non-statistically significant correlations were found ranging from no correlation to $r = -0.39$ between motion parameters and patient-rated outcomes (Table 3 A,B and C). The 2D circumduction motion parameter was the only lumbar motion parameter significantly correlated with pain score and RMDQ score in the total cohort ($p < .01$), but the correlation coefficients were low, ranging from -0.20 to -0.22 (Table 3A). For most motion parameters stronger correlations were found in the group with back pain only (motion parameters vs. pain intensity ranging from -0.37 to 0.05) compared to the group with back and leg pain (motion parameters vs. pain intensity ranging from -0.11 to 0.08), and the motion parameters showing the strongest correlations to patient-rated outcomes were phase-plot area, sagittal ROM and 2D circumduction motion (Table 3B). In addition stronger correlations were found in the SMT group (motion parameters vs. pain intensity ranging from -0.39 to 0.03) (Table 3C).

Our hypothesis that patients who had a clinically relevant improvement in pain and back-related function would have greater change scores in velocity, ROM, circumduction area, and have a smoother motion compared to patients who did not achieve a clinically relevant improvement, could also not be confirmed (Table 4). Some motion parameters however did change as hypothesized but the confidence intervals were wide i.e. none were significant (Table 4). In general, the participants who deteriorated, i.e. experienced at least 30% deterioration in back pain, decreased in all motion parameters and changed to a less smooth motion. However the same consistent patterns were not found between motion parameters and back-related function and in addition only 8 participants deteriorated.

For sub-groups based on pain distribution, differences were identified in several but not all motion parameters when comparing changes for participants with back pain only versus back pain and leg pain (Table 5). For patients with back pain only, all motion parameters changed in the hypothesized

direction i.e. the motion parameter scores increased when clinically relevant changes were reported both relative to back pain intensity and to RMDQ score, with the exception of the Jerk Index. For the participants with back pain only and no clinically relevant improvement, zero was included in the 95% confidence intervals. This was different to those who had a clinically relevant improvement, with the exception of the Jerk Index and the median confidence intervals for the maximum flexion velocity parameter. Significant differences between the clinically relevant improvement and no clinical relevant change were found for the motion parameters ROM (5.9 (95% CI 2.3-9.4) vs. 1.0 (95% CI -3.7-5.7) degree) and phase-plot area (993 (95% CI 711-1182) vs. 161 (95% CI -558-828) degree² s⁻¹) for back pain intensity in the back pain only subgroup. Significant increase in ROM (5.3 (95% CI 1.9-8.8) vs. -0.4 (95% CI -5.4-4.6) degree) was found between clinically relevant improvement and no clinical relevant change groups for back-related function (Table 5). This was in contrast to the group with back and leg pain, where several motion parameters changed in the opposite direction to that hypothesized and no significant differences was found.

Finally we assessed the relationship between treatment on pain and back-related function versus motion parameters (Table 6). The SET group experienced the highest percentage of clinically relevant improvement compared to the other groups i.e. 85% for back pain and 77% for RMDQ. This was followed closely by the other two groups that ranged from 60% to 70%. In general, the exercise groups had higher lumbar motion change scores than the SMT group when no clinically relevant changes were reported (Table 6). Significant increase between the clinically relevant improvement and no clinical relevant change was found for the motion parameter phase-plot area in the SMT group (back pain 978 (95% CI 458-1309) vs. -305 (95% CI -880-858)) and (RMDQ 1054 (95% CI 467-1374) vs. 267 (95% CI -394-864)) (Table 6). The relative difference in the regional lumbar motion change scores between

clinically relevant improved versus no clinical relevant change was consistently higher for the SMT group (Table 6) and were stronger correlated than the exercise groups (Table 3 C).

Discussion

In this study, we assessed the relationship between change scores in patient-rated outcomes (pain and back-related function) and several objective measures of regional lumbar motion in a group of chronic LBP patients. In general, the changes in regional lumbar motion parameters investigated in this study do not tell us much about the changes in patients' perceived back-related function or back pain intensity and therefore, as proposed by others, may not be good proxy measures of patient-rated outcomes^{4;15;20}. However, this study provides novel evidence that the distribution of back/leg pain and treatment received, affect associations between motion parameters and patient-rated outcomes. For instance, patients receiving 'active' treatment in the form of exercise therapy changed in regional lumbar motion regardless of the clinical relevant improvement, whereas patients receiving the more 'passive' therapy of spinal manipulation regional lumbar motion changed only in patients experiencing a clinical relevant improvement (Table 3C and Table 6).

The way in which a person uses his/her back may determine the presence or absence of pain³¹ and a potential attraction of kinematic assessments is the notion that they might display abnormalities reflective of an underlying disease⁵. For instance, a person suffering from back pain may avoid certain postures that cause pain, or similarly, muscle activation patterns may be altered because of pain. Therefore, a functional kinematic assessment might seem to be a logical choice in order to differentiate between subtypes of back pain or evaluate progress over time. However, the actual usefulness of regional lumbar motion measurements remains controversial, especially the uncertain relationship between patient perception and what can be measured objectively. Lumbar motion measurements are

probably influenced by several subjective factors such as the patient's agenda, motivation, effort, fear and other psychosocial states, as well as actual physical capabilities ^{4;32-35} and these biopsychosocial factors may also be related to the presence of chronic LBP ³⁶. Cox et al. stated that the use of more refined measurements that are relatively independent of patient control may offer a better representation of 'true' spinal dysfunction ³²; however as commented on by McGregor et al., such measurements are not always feasible in the clinical environment⁴.

The correlations between subjective (patient-rated) and objective (regional lumbar motion) change values were generally low or non-existent. These weak associations between changes in objective and subjective back-related function have also been found by other researchers ^{4;6;19;20}, and our results additionally show that the kinematic response to treatment is less predictable when leg pain is also present.

We hypothesized that removing the highest baseline quartile of the motion parameters would result in a stronger correlation because of a ceiling effect in the motion parameters e.g. patients presenting with high mobility might not be able to improve in their regional lumbar motion. Indeed, we found that the highest baseline quartile changed differently when compared to the other quartiles (Table 2) but surprisingly, removing these participants from the analysis did not result in stronger correlations overall and did not change other conclusions. Choosing other cut-points may have resulted in slightly different values but they were unlikely to have changed the overall picture. Another consideration is that for the individual patient, stratification based on a certain baseline score may be inappropriate because some patients with LBP may not have a functional limitation relative to what is normal for that particular individual but may still have limitations when compared to other patients. The changes in the two outer quartiles could be explained by the regression toward the mean phenomenon.

There are several limitations in the current study that need to be taken into consideration.

The fixed extremity of the CA6000 Spine Motion Analyzer was mounted on the sacral crest (S2) and the mobile end of the device was mounted at the level of T7 spinous process. Therefore the measured movements were due both lumbar motion and motion from the lower part of the thoracic region.

Only mean pain intensity score from the past week was collected for this RCT and, because of the fluctuation of the condition, it may have been more appropriate to have used a pain score at the exact time point of the lumbar motion assessment. However, we have no data on pain during the testing procedure or the potential influence this might have had on the results. However, all patients who were evaluated were able to complete the lumbar motion test trials.

We found that the participants with the highest baseline motion scores changed differently from the rest of the sample, indicating that a considerable proportion of the study sample may have displayed a ceiling effect, which might question the use of this tool for measuring changes in this population. The directions of the changes in the upper and lower quartiles indicate a regression toward the mean phenomenon (Table 2).

We did include patients who became worse in the analysis and data is presented in tables but interpretation of data should be done with caution, firstly, because the groups were very small (ranging from 1 to 8 participants), and secondly, because improvement and deterioration may be different concepts e.g. larger changes may be needed for patients to feel worse than to feel better³⁷.

In any assessment, measurements should be reliable and valid for the physiology being measured; however several factors affect reliability and measurement error, especially when using instruments attached to the skin surface. Additional measurement error is likely to be present when the instrument is attached on a patient wearing a t-shirt as done in this setup. However in general, the reliability of this

test procedure would be considered to be moderate based on the ICCs and therefore useful for analysis at a group level but not on the individual patient level^{24;25;38}. In addition the variation in ICCs (ranging from 0.55-0.81) indicated that some motion parameters were more reliable than others. Moreover, only 199 patient recordings of the 301 included patients were available for motion analysis due to technical problems and dropouts. Even though our analysis did not indicate any major differences between participants and non-participants in this analysis, we cannot rule out potential bias caused by missing data. Collectively, these limitations may have affected the results of this study. Future studies should examine relationships between lumbar motion versus pain and back-related function measures just before and during movement test procedure instead of past week mean pain level as assessed in the present study.

Conclusion

Overall, changes in regional lumbar motion were poorly associated with back pain intensity scores measured by ordinal 11-box scale for the previous week and back-related function measured by RMDQ. This could not be explained by a ceiling effect. However, associations between regional lumbar motion versus patient-rated pain and back-related function were different relative to subgroups. Thus stronger correlation coefficients and significant differences between clinically relevant improved versus no clinical relevant change were found in some motion parameters in the subgroup with back pain only and the treatment group receiving spinal manipulation.

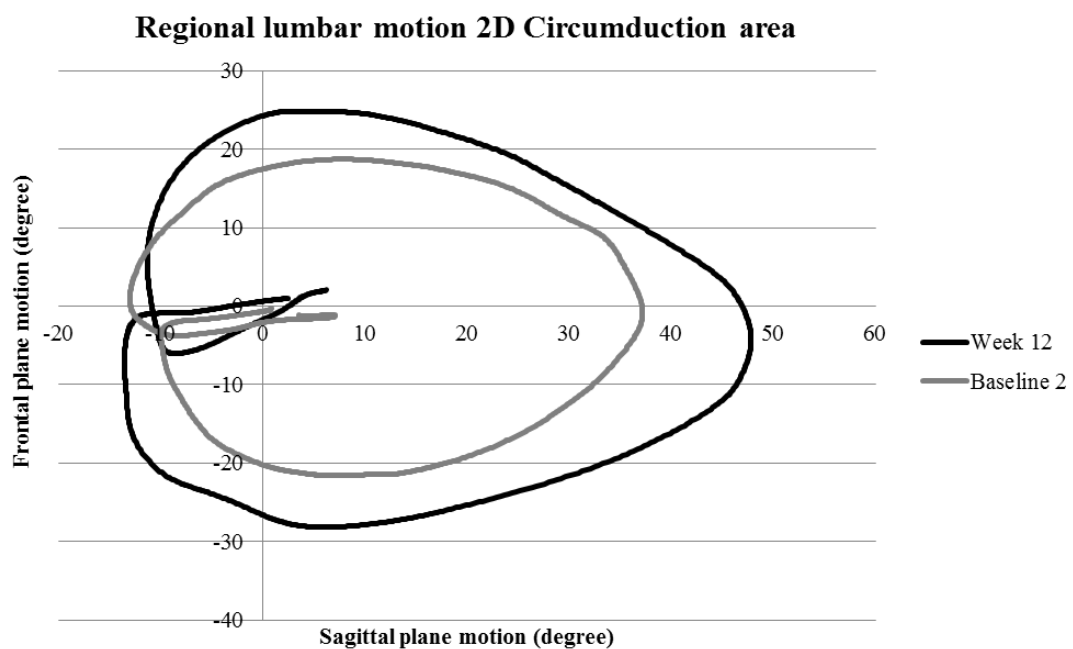
References

1. Ferguson SA, Marras WS, Gupta P. Longitudinal quantitative measures of the natural course of low back pain recovery. *Spine* 2000 1;25(15):1950-6.
2. Henderson CN. The basis for spinal manipulation: Chiropractic perspective of indications and theory. *J Electromyogr Kinesiol* 2012 16;22(5):632-42.
3. Kang SW, Lee WN, Moon JH, Chun SI. Correlation of spinal mobility with the severity of chronic lower back pain. *Yonsei Med J* 1995;36(1):37-44.
4. McGregor AH, Hughes SPF. The potential use of spinal motion as a measure of surgical outcome. *J Back Musculoskelet* 2004;17(2):77-82.
5. Bishop JB, Szpalski M, Ananthraman SK, McIntyre DR, Pope MH. Classification of low back pain from dynamic motion characteristics using an artificial neural network. *Spine* 1997 15;22(24):2991-8.
6. Burton AK, Tillotson KM, Edwards VA, Sykes DA. Lumbar sagittal mobility and low back symptoms in patients treated with manipulation. *J Spinal Disord* 1990;3(3):262-8.
7. Cherniack M, Dillon C, Erdil M, Ferguson S, Kaplan J, Krompinger J, et al. Clinical and psychological correlates of lumbar motion abnormalities in low back disorders. *Spine J* 2001;1(4):290-8.
8. Dickey JP, Pierrynowski MR, Bednar DA, Yang SX. Relationship between pain and vertebral motion in chronic low-back pain subjects. *Clin Biomech* 2002;17(5):345-52.
9. Gronblad M, Hurri H, Kouri JP. Relationships between spinal mobility, physical performance tests, pain intensity and disability assessments in chronic low back pain patients. *Scand J Rehabil Med* 1997;29(1):17-24.
10. Hsieh CY, Pringle RK. Range of motion of the lumbar spine required for four activities of daily living. *J Manipulative Physiol Ther* 1994;17(6):353-8.
11. Lehman GJ. Biomechanical assessments of lumbar spinal function. How low back pain sufferers differ from normals. Implications for outcome measures research. Part I: kinematic assessments of lumbar function. *J Manipulative Physiol Ther* 2004;27(1):57-62.
12. McGregor AH, Cattermole HR, Hughes SPF. Global spinal motion in subjects with lumbar spondylolysis and spondylolisthesis - Does the grade or type of slip affect global spinal motion? *Spine* 2001 1;26(3):282-6.
13. Moutzouri M, Billis E, Strimpakos N, Kottika P, Oldham JA. The effects of the Mulligan Sustained Natural Apophyseal Glide (SNAG) mobilisation in the lumbar flexion range of asymptomatic subjects as measured by the Zebris CMS20 3-D motion analysis system. *BMC Musculoskel Dis* 2008;9:131.
14. Nattrass CL, Nitschke JE, Disler PB, Chou MJ, Ooi KT. Lumbar spine range of motion as a measure of physical and functional impairment: an investigation of validity. *Clin Rehabil* 1999;13(3):211-8.

15. Sullivan MS, Shoaf LD, Riddle DL. The relationship of lumbar flexion to disability in patients with low back pain. *Phys Ther* 2000;80(3):240-50.
16. Zuberbier OA, Kozlowski AJ, Hunt DG, Berkowitz J, Schultz IZ, Crook JM, et al. Analysis of the convergent and discriminant validity of published lumbar flexion, extension, and lateral flexion scores. *Spine* 2001 15;26(20):E472-E478.
17. Magnusson ML, Bishop JB, Hasselquist L, Spratt KF, Szpalski M, Pope MH. Range of motion and motion patterns in patients with low back pain before and after rehabilitation. *Spine* 1998 1;23(23):2631-9.
18. Marras WS, Parnianpour M, Ferguson SA, Kim JY, Crowell RR, Bose S, et al. The Classification of Anatomic-Based and Symptom-Based Low-Back Disorders Using Motion Measure Models. *Spine* 1995 1;20(23):2531-46.
19. Taimela S, Harkapaa K. Strength, mobility, their changes, and pain reduction in active functional restoration for chronic low back disorders. *J Spinal Disord* 1996;9(4):306-12.
20. Poitras S, Loisel P, Prince F, Lemaire J. Disability measurement in persons with back pain: A validity study of spinal range of motion and velocity. *Arch Phys Med Rehab* 2000;81(10):1394-400.
21. Bronfort G, Maiers MJ, Evans RL, Schulz CA, Bracha Y, Svendsen KH, et al. Supervised exercise, spinal manipulation, and home exercise for chronic low back pain: a randomized clinical trial. *Spine J* 2011;11(7):585-98.
22. Jensen MP, Karoly P, Braver S. The measurement of clinical pain intensity: a comparison of six methods. *Pain* 1986;27(1):117-26.
23. Patrick DL, Deyo RA, Atlas SJ, Singer DE, Chapin A, Keller RB. Assessing Health-Related Quality-Of-Life in Patients with Sciatica. *Spine* 1995 1;20(17):1899-908.
24. Portney LG, Watkins M.P. *Foundation of clinical research: applications to practice*. Third ed. Upper Saddle River, New Jersey: Pearson Education, Inc.; 2009.
25. Mieritz RM, Bronfort G, Kawchuk G, Breen A., Hartvigsen J. Reliability and measurement error of 3D regional lumbar motion measures: a systematic review. *J Manipulative Physiol Ther* 2012;35(8):645-56.
26. Mannion A, Troke M. A comparison of two motion analysis devices used in the measurement of lumbar spinal mobility. *Clin Biomech* 1999;14(9):612-9.
27. Ostelo RWJG, Deyo RA, Stratford P, Waddell G, Croft P, Von Korff M, et al. Interpreting change scores for pain and functional status in low back pain - Towards international consensus regarding minimal important change. *Spine* 2008 1;33(1):90-4.
28. Park W, Ramachandran J, Weisman P, Jung ES. Obesity effect on male active joint range of motion. *Ergonomics* 2010;53(1):102-8.
29. Troke M, Moore AP, Maillardet FJ, Cheek E. A normative database of lumbar spine ranges of motion. *Man Ther* 2005;10(3):198-206.

30. WHO Expert Committee. Physical status: The use and interpretation of anthropometry. Geneva: WHO; 1995.
31. Adams MA. Biomechanics of back pain. *Acupunct Med* 2004;22(4):178-88.
32. Cox ME, Asselin S, Gracovetsky SA, Richards MP, Newman NM, Karakusevic V, et al. Relationship between functional evaluation measures and self-assessment in nonacute low back pain. *Spine* 2000 15;25(14):1817-26.
33. Geisser ME, Haig AJ, Wallbom AS, Wiggert EA. Pain-related fear, lumbar flexion, and dynamic EMG among persons with chronic musculoskeletal low back pain. *Clinical Journal of Pain* 2004;20(2):61-9.
34. Marras WS, Davis KG, Heaney CA, Maronitis AB, Allread WG. The influence of psychosocial stress, gender, and personality on mechanical loading of the lumbar spine. *Spine* 2000 1;25(23):3045-54.
35. Thomas JS, France CR, Lavender SA, Johnson MR. Effects of fear of movement on spine velocity and acceleration after recovery from low back pain. *Spine* 2008 1;33(5):564-70.
36. Langevin HM, Sherman KJ. Pathophysiological model for chronic low back pain integrating connective tissue and nervous system mechanisms. *Med Hypotheses* 2007;68(1):74-80.
37. De Vet HCW, Terwee CB, Ostelo RW, Beckerman H, Knol DL, Bouter LM. Minimal changes in health status questionnaires: distinction between minimally detectable change and minimally important change. *Health Qual Life Outcomes* 2006;4:54.
38. Aaronson N, Alonso J, Burnam A, Lohr KN, Patrick DL, Perrin E, et al. Assessing health status and quality-of-life instruments: Attributes and review criteria. *Qual Life Res* 2002;11(3):193-205.

Figure 1



Clockwise circumduction motion in a typical patient before and after treatment. The area increased after treatment.

Table 1

Baseline characteristics for 199 chronic LBP patients participating in a randomized clinical trial

Intervention group (n)	SMT (77)		SET (62)		HEA (60)		ALL (199)		NMD (80)	
Characteristic (n=199)	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Age (years)	46	11	45	12	47	11	46	11	44*	10
BMI	28	6	28	6	28	5	28	6	28	5
Males (%)	35		45		40		40		40	
<u>Back pain level (0-10)</u>										
Baseline 1	5.3	1.5	5	1.4	5.1	1.6	5.1	1.5	5.5	1.7
Baseline 2	5.3	1.8	5	1.5	5.1	1.8	5.2	1.7	5.2	1.7
Pain more than 1 year (%)	84		77		83		82		83	
<u>Quebec Task Force Classification (%)</u>										
	1	69	66		67		67		67	
	2	19	18		17		18		18	
	3	12	15		12		13		10	
	4		2		5		2		5	
SF36 Physical Component	43	7	44	7	43	8	43	7	42	8
SF36 Mental Component	55	8	54	8	54	8	54	8	54	8
Depression (CESD)	10	10	13	11	11	9	11	10	13	10
RMDQ	36	20	36	21	37	24	36	21	37	21

(SMT) = Spinal manipulative treatment, (SET) = Supervised exercise therapy and (HEA) = Home exercise and advice, (NMD) = no motion data due to instrument error, * = Individuals who did not complete (n=80) were significantly different to those who did (n=199), Depression is defined as greater than 16 points on the Center for Epidemiologic Studies Depression Scale (CESD), BMI = Body Mass Index (kg/m²), RMDQ = Roland Morris Disability Questionnaire.

Table 2

Lumbar motion changes stratified in baseline score quartiles

	100-75 quartile			75-50 quartile			50-25 quartile			25-0 quartile		
Measure	n	Median (95% CI)		n	Median (95% CI)		n	Median (95% CI)		n	Median (95% CI)	
P-P	50	-514 (-1284-87)		* 50	501 (-273-869)		50	795 (437-1014)		49	800 (573-1077)	
Vel	50	-6.2 (-11.0-3.3)		* 49	1.6 (.3.6-6.7)		51	2.0 (-0.2-3.3)		49	6.9 (4.3-10.0)	
Jerk	50	1.3 (0.7-2.1)		49	-0.1 (-0.9-1.3)		50	0.8 (-1.0-2.0)		50	-3.3 (-3.9--2.0)	
#ROM	50	-6.3 (-9.7--2.8)		* 49	6.3 (2.0-10.5)		51	6.8 (3.7-9.8)		49	9.8 (6.5-13.2)	
2D	50	-124 (-410-229)		49	118 (-214-466)		51	193 (-94-536)		49	375 (207-522)	
3D	49	-29.6 (-67.4-3.0)		* 51	16.5 (-1.9-39.1)		49	32.8 (11.3-52.1)		50	30.1 (18.8-45.7)	

P-P =Phase-plot Area ($^{\circ}$ *($^{\circ}$ /sec)), Vel = Maximum flexion velocity ($^{\circ}$ /sec), Jerk = Jerk index (number of changes in acceleration from full extension to full flexion), ROM= Range of motion ($^{\circ}$), 2D = 2-dimensional circumduction area (degree²), 3D = 3-dimensional circumduction area (cm²), # = Mean calculated instead of medians because ROM was normally distributed, * = Significant difference between quartiles based on 95% CI intervals, NB: the Jerk Index is displayed in an inversed format because our hypothesis of improvement was oposite i.e. people who already moved smoothly (low index) had little capacity to improve further.

Table 3

Rank -order correlation coefficients (Spearman r) between Back pain intensity (BPI) or RMDQ reduction and motion parameter changes following 12-week interventions

A) Stratified by quartile ((a)= all patients available, (b) = highest motion scores (75% percentile) from the second baseline measurement were not included)

All participants

	BPI (a)	BPI (b)	RMDQ (a)	RMDQ (b)
n	199	144-141	199	144-141
P-P	-0.12	-0.16	-0.13	-0.21
Vel	-0.06	0.00	-0.03	-0.03
Jerk	0.02	-0.02	-0.03	-0.13
ROM	-0.09	-0.12	-0.15	-0.24*
2D	-0.22*	-0.17	-0.20*	-0.16
3D	-0.09	0.00	-0.14	-0.06

B) Stratified by pain distribution and quartile ((a)= all patients available, (b) = highest motion scores (75% percentile) from the second baseline measurement were not included)

Back pain only					Back pain and Leg pain			
	BPI (a)	BPI (b)	RMDQ (a)	RMDQ (b)	BPI (a)	BPI (b)	RMDQ (a)	RMDQ (b)
n	89	60-64	89	60-64	110	78-82	110	78-82
P-P	-0.36*	-0.24	-0.27	-0.17	0.05	-0.11	-0.02	-0.25
Vel	-0.23	-0.07	-0.16	-0.09	0.08	0.03	0.07	0.00
Jerk	0.05	-0.01	-0.07	-0.13	0.00	-0.05	0.01	-0.13
ROM	-0.22	-0.20	-0.14	-0.16	0.01	-0.07	-0.15	-0.30*
2D	-0.37*	-0.34*	-0.30*	-0.26	-0.11	-0.07	-0.12	-0.10
3D	-0.14	-0.07	-0.11	-0.02	-0.07	0.04	-0.17	-0.10

C) Stratified by treatment groups

Measure	SET		SMT		HEA	
	BPI	RMDQ	BPI	RMDQ	BPI	RMDQ
n	62	62	77	77	60	60
P-P	0.11	0.07	-0.39*	-0.34*	-0.01	-0.02
Vel	0.09	0.13	-0.21	-0.15	-0.07	-0.05
Jerk	0.10	-0.11	0.03	0.03	0.05	0.06
ROM	-0.02	-0.02	-0.28	-0.24	0.05	-0.14
2D	-0.07	-0.14	-0.32*	-0.21	-0.18	-0.20
3D	0.01	-0.02	-0.20	-0.21	0.01	-0.12

P-P = Phase-plot Area ($^{\circ}$ *($^{\circ}$ /sec)), Vel = Maximum flexion velocity ($^{\circ}$ /sec), Jerk = Jerk index (number of changes in acceleration from full extension to full flexion), ROM = Range of motion ($^{\circ}$), 2D = 2-dimensional circumduction area (degree²), 3D = 3-dimensional circumduction area (cm²), BPI = Back pain intensity, RMDQ = Roland Morris Disability Questionnaire, * = p < 0.01, SET = Supervised exercise therapy, SMT = Spinal manipulative therapy, HEA = Home exercise and advice.

Table 4

Regional lumbar motion changes vs. clinically relevant changes in all included patients

All	>30% Improvement in BPI (n=143)	no change in BPI (n=48)	>30% Deterioration in BPI (n=8)
Measure	Median (95% CI)	Median (95% CI)	Median (95% CI)
P-P	580 (330-850)	296 (-206-831)	-675 (-1349-2277)
Vel	2,3 (-0.2-4.3)	2,8 (-0.5-6.1)	-2,0 (-13.9-9.5)
Jerk	0,05 (-0.7-0.8)	-0,1 (-1.2-1.1)	1,2 (-3.2-9.5)
#ROM	4,8 (2.5-7.1)	2,8 (-0.9-6.6)	-0,7 (-15.2-13.8)
2D	210 (108-388)	197 (-110-365)	-402 (-1335-934)
3D	18 (8-29)	28 (7-48)	-20 (-112-74)

All	>30% Improvement in RMDQ (n=136)	no change in RMDQ (n=53)	>30% Deterioration in RMDQ (n=8)
Measure	Median (95% CI)	Median (95% CI)	Median (95% CI)
P-P	603 (314-845)	308 (-240-830)	462 (-1109-1237)
Vel	2,0 (-0.5-5.0)	2,0 (0.6-4.1)	-0,2 (-9.2-8.0)
Jerk	0,2 (-0.7-0.9)	-0,5 (-1.2-0.8)	0,4 (3.8-2.5)
#ROM	5,1 (2.8-7.5)	1,8 (-1.8-5.4)	2,5 (-6.8-11.8)
2D	246 (108-402)	24 (-255-334)	164 (-901-474)
3D	21 (9-31)	11 (-14-33)	22 (-70-51)

P-P =Phase-plot Area ($^{\circ}$ *($^{\circ}$ /sec)), Vel = Maximum flexion velocity ($^{\circ}$ /sec),
 Jerk = Jerk index (number of changes in acceleration from full extension to full flexion), ROM= Range of motion ($^{\circ}$), 2D = 2-dimensional circumduction area (degree²), 3D = 3-dimensional circumduction area (cm²),
 RMDQ = Roland Morris Disability Questionnaire, BPI =back pain intensity,
 # Mean calculated instead of medians because ROM was normally distributed.

Table 5

Regional lumbar motion changes vs. clinically relevant changes stratified by LBP patients with and without leg pain

Patients with back pain only

	>30% Improvement in BPI (n=58)		no change in BPI (n= 27)	>30% Deterioration in BPI (n=4)
Measure	Median (95% CI)		Median (95% CI)	Median (95% CI)
P-P	993 (711-1182)	*	161 (-558-828)	-1077 (-1624--295§)
Vel	4,0 (-0.2-6.9)		0,7 (-5-6.5)	-2,0 (-11.8-3.4§)
Jerk	-0,2 (-1.7-0.9)		0,2 (-1.9-1.5)	0,8 (-7-5.8§)
#ROM	5,9 (2.3-9.4)	*	1,0 (-3.7-5.7)	-11 (-29-5.9)
2D	289 (118-559)		136 (-239-371)	-722 (-1821-24§)
3D	31 (4.5-50)		29 (-2.9-50)	20 (-160-22§)
	>30% Improvement in RMDQ (n=60)		no change in RMDQ (n=25)	>30% Deterioration in RMDQ (n=4)
Measure	Median (95% CI)		Median (95% CI)	Median (95% CI)
P-P	844 (463-1161)		267 (-398-905)	593 (-2065-1124§)
Vel	3,4 (-0.3-6.7)		3,0 (-3.9-5.3)	1,9 (-5.5-8§)
Jerk	0,02 (-1.3-1.4)		-1,1 (-2.5-0.7)	-1,2 (-5-1.5§)
#ROM	5,3 (1.9-8.8)	*	-0,4 (-5.4-4.6)	2,5 (-30-35)
2D	282 (86-528)		24 (-410-401)	136 (-1319-369§)
3D	30 (4.2-50)		17 (-17-47)	36 (-120-52§)

Patients with back and leg pain

	>30% Improvement in BPI (n=85)		no change in BPI (n=21)	>30% Deterioration in BPI (n=4)
Measure	Median (95% CI)		Median (95% CI)	Median (95% CI)
P-P	316 (14-581)		666 (-241-1218)	1081 (-1217-4099§)
Vel	1,4 (-2.3-3.5)		3,3 (-0.7-9.7)	2,8 (-18-22§)
Jerk	0,6 (-0.7-1.1)		-0,5 (-1.9-1.4)	5,9 (-1.3-10.6§)
#ROM	4,1 (1.0-7.5)		5,3 (-1.2-11.7)	10,0 (-16.2-36.3)
2D	138 (-97-384)		265 (-238-633)	134 (-1101-1783§)
3D	15 (4-24)		16 (-1.5-48)	8 (-89-137§)
	>30% Improvement in RMDQ (n=76)		no change in RMDQ (n=28)	>30% Deterioration in RMDQ (n=6)
Measure	Median (95% CI)		Median (95% CI)	Median (95% CI)
P-P	391 (56-698)		397 (-294-912)	342 (-1184-2009)
Vel	0,3 (-2.5-4.4)		2,0 (0.01-5.6)	0,8 (-17.3-20.9)
Jerk	0,7 (-0.8-1.2)		-0,4 (-1.1-1.4)	0,4 (-3.9-8.3)
#ROM	5,0 (1.6-8.4)		3,7 (-1.6-9.0)	2,5 (-5.1-10.1)
2D	225 (-58-426)		4 (-253-420)	236 (-1040-733)
3D	18 (6-29)		6 (-25-32)	18 (-83-49)

P-P = Phase-plot Area (**(°/sec)), Vel = Maximum flexion velocity (°/sec), Jerk = Jerk index (number of changes in acceleration from full extension to full flexion), ROM = Range of motion (°), 2D = 2-dimensional circumduction area (degree²), 3D = 3-dimensional circumduction area (cm²), RMDQ = Roland Morris Disability Questionnaire, BPI = back pain intensity, # = Mean calculated instead of median because ROM was normally distributed. * = Significant difference between >30% improved and no change groups based on 95% CI intervals, § = Lower (upper) confidence limit held at minimum (maximum) of sample.

Table 6

Regional lumbar motion changes vs. clinical relevant changes in back pain or RMDQ by treatment group

Supervised exercise therapy (SET)

	>30% Improvement in BPI (n=53)	no change in BPI (n=7)	>30% Deterioration in BPI (n=2)
Measure	Median (95% CI)	Median (95% CI)	Median (95% CI)
P-P	316 (-32-710)	817 (-783-2060)	-169 (-1101-763§)
Vel	0,0 (-2.6-3.4)	8,7 (-3.2-21.6)	-14,9 (-18.1--11.8§)
Jerk	1,2 (-0.3-1.6)	0,2 (-4.2-6.9)	7,4 (5.8-8.9§)
#ROM	3,2 (-0.8-7.1)	2,5 (-5.2-10.2)	-1,5 (-109-105)
2D	181 (-1.0-441)	300 (-383-959)	-186 (-897-525§)
3D	18 (4-30)	48 (-16-106)	-58 (-160-44§)
	>30% Improvement in RMDQ (n=48)	no change in RMDQ (n=12)	>30% Deterioration in RMDQ (n=2)
Measure	Median (95% CI)	Median (95% CI)	Median (95% CI)
P-P	297 (-44-805)	249 (-896-1088)	1426 (763-2089§)
Vel	-0,2 (-2.7-4.3)	1,5 (-5-13.6)	2,3 (-18-22.6§)
Jerk	1,2 (-0.3-2.5)	1,0 (-2.3-1.6)	2,5 (-4.0-8.9§)
#ROM	3,1 (-1.1-7.3)	1,8 (-5.0-8.6)	1,8 (-5.0-8.6)
2D	283 (79-521)	-55 (-642-592)	359 (193-525§)
3D	16 (3-29)	32 (-22-60)	46 (44-49§)

Spinal manipulative therapy (SMT)

	>30% Improvement in BPI (n=52)	no change in BPI (n=20)	>30% Deterioration in BPI (n=5)
Measure	Median (95% CI)	Median (95% CI)	Median (95% CI)
P-P	978 (458-1309) *	-305 (-880-858)	-1054 (-1624-4099§)
Vel	4,7 (0.5-7.1)	1,4 (-7.8-5.0)	-0,7 (-5.1-22.1§)
Jerk	-1,4 (-2.4--0.1)	-1,0 (-2.6-1.0)	-0,4 (-1.3-10.6§)
#ROM	6,4 (2.7-10.2)	0,4 (-6.9-7.7)	-9,9 (-28-27)
2D	265 (-35-568)	-2 (-319-411)	-547,7 (-1820-1782§)
3D	23 (-1-45)	12 (-18-47)	-22,2 (-89-136§)
	>30% Improvement in RMDQ (n=46)	no change in RMDQ (n=27)	>30% Deterioration in RMDQ (n=4)
Measure	Median (95% CI)	Median (95% CI)	Median (95% CI)
P-P	1054 (467-1374) *	267 (-394-864)	-1051 (-2065-1292§)
Vel	6,0 (-2.3-8.1)	1,9 (-0.3-4.7)	-5,3 (-11.0-5.2§)
Jerk	-0,8 (-2.2-0.4)	-1,2 (-2.3--0.4)	-1,3 (-5.1-2.9§)
#ROM	6,7 (2.5-10.8)	2,0 (-4.0-8.0)	2,0 (-4.0-8.0)
2D	242 (-44-544)	-207 (-293-435)	-411 (-1319-756§)
3D	35 (3.3-65)	4 (-22-18)	-39 (-120-25§)

Home exercise and advice (HEA)

	>30% Improvement in BPI (n=38)	no change in BPI (n=21)	>30% Deterioration in BPI (n=1)
Measure	Median (95% CI)	Median (95% CI)	Median (95% CI)
P-P	521 (129-940)	572 (96-1016)	-295
Vel	2,9 (-3.9-6.9)	2,5 (-2.6-8.5)	3,4
Jerk	0,6 (-1.9-1.4)	0,9 (-1.0-1.4)	-7
#ROM	4,9 (0.3-9.6)	5,3 (0.1-10.5)	0,3
2D	130 (-136-505)	265 (-192-389)	24
3D	15 (-13-31)	33 (2-55)	22
	>30% Improvement in RMDQ (n=42)	no change in RMDQ (n=14)	>30% Deterioration in RMDQ (n=4)
Measure	Median (95% CI)	Median (95% CI)	Median (95% CI)
P-P	444 (210-828)	746 (-527-1533)	593 (-80-1124§)
Vel	1,5 (-1.2-5.8)	3,4 (-8-9)	5,6 (-3.8-8.1§)
Jerk	0,4 (-1.8-1.3)	0,8 (-1.7-2.5)	0,9 (-3.4-1.5§)
#ROM	5,9 (1.5-10.2)	1,4 (-5.0-7.8)	1,4 (-5.0-7.8)
2D	230 (-34.8-518)	68 (-337-412)	136 (-483-369§)
3D	24 (8-33)	16 (-19-64)	36 (-30-52§)

P-P = Phase-plot Area (**(°/sec)), Vel = Maximum flexion velocity (°/sec), Jerk = Jerk index (number of changes in acceleration from full extension to full flexion), ROM = Range of motion (°), 2D = 2-dimensional circumduction area (degree²), 3D = 3-dimensional circumduction area (cm²), BPI = back pain intensity, RMDQ = Roland Morris Disability Questionnaire, r = spearman correlation coefficient, * = Significant difference between >30% improved and no change groups based on 95% CI intervals, § = Lower (upper) confidence limit held at minimum (maximum) of sample, # = Mean calculated instead of meadian because ROM was normally distributed.